

# Stable reconstructions for the analysis formulation of $\ell^p$ -minimization using redundant systems

Jackie Ma\*

Technische Universität Berlin  
Department of Mathematics  
Straße des 17. Juni 136, 10623 Berlin

September 21, 2016

## Abstract

In compressed sensing sparse solutions are usually obtained by solving an  $\ell^1$ -minimization problem. Furthermore, the sparsity of the signal does not need to be directly given. In fact, it is sufficient to have a signal that is sparse after an application of a suitable transform. In this paper we consider the stability of solutions obtained from  $\ell^p$ -minimization for arbitrary  $0 < p \leq 1$ . Further we suppose that the signals are sparse with respect to general redundant transforms associated to not necessarily tight frames. Since we are considering general frames the role of the dual frame has to be additionally discussed. For our stability analysis we will introduce a new concept of so-called *frames with identifiable duals*. Further, we numerically highlight a gap between the theory and the applications of compressed sensing for some specific redundant transforms.

**Keywords** Compressed sensing, redundant transforms, non-convex, restricted isometry property, wavelets, shearlets

**Mathematics Subject Classification** 42C15 41A65 46C05 42C40

## 1 Introduction

Suppose we are interested in the reconstruction of a certain object of interest but we are not enabled to observe the signal directly. We are instead allowed to sense it. Mathematically speaking, we understand this problem as solving a system of linear equations

$$Ax = y, \tag{1.1}$$

where  $x \in \mathbb{R}^n$  represents the object of interest,  $A \in \mathbb{R}^{m \times n}$  is the sensing matrix and  $y \in \mathbb{R}^m$  represents the resulting observations. It is of course preferred to keep the amount of acquired data small, i.e.  $m$  should be small. However, if  $m$  is (possibly) much smaller than  $n$ , then (1.1) usually does not have a unique solution in general. This issue can be resolved by imposing additional assumptions. One of such possible assumptions is the concept of *sparsity*. We say a signal  $x$  is *sparse*, if the vector  $x$  contains only very few non-zero entries. This concept of sparsity is fundamental in the field of *compressed sensing* [8, 11, 17] where one aims to recover sparse signals using only very few measurements. With a view to solve (1.1) a *compressed sensing reconstruction* is then typically computed by solving an *equality constrained minimization* problem of the form

$$\min_x f(x) \quad \text{subject to} \quad y = Ax, \tag{P}$$

---

\*ma@math.tu-berlin.de

where  $f : \mathbb{R}^n \rightarrow \mathbb{R}^+$  is some non-negative function, the so-called *prior-function* or *regularizer*. However, it is often not feasible in practice to acquire the data  $y$  without any errors or noise this means instead of having (1.1) we rather have

$$Ax \approx y$$

where we use  $a \approx b$  to indicate  $a$  is close to  $b$  but not necessarily equal. Therefore, in order to allow perturbations during the measurement process, the model (P) is extended to

$$\min_x f(x) \quad \text{subject to} \quad \|y - Ax\|_2 \leq \varepsilon, \quad (\text{P}^\varepsilon)$$

where  $\varepsilon > 0$  is a control parameter for the error that may arise during the acquisition. Note that any element satisfying the constraint in (P) also satisfies the constraint in  $(\text{P}^\varepsilon)$ .

A major question is of course how to choose the prior  $f : \mathbb{R}^n \rightarrow \mathbb{R}^+$ . In the spirit of compressed sensing the ideal choice would be

$$f(x) = \|x\|_0 := \#\{k : x_k \neq 0\}$$

since for this particular choice the problem (P) returns the sparsest solution matching the constrained. Unfortunately, for  $f(x) = \|x\|_0$  the  $\ell^0$ -problem (P) is NP-hard [22]. A common escape is a convex relaxation of the  $\ell^0$ -problem by using, for example

$$f(x) = \|x\|_1 = \sum_{k \leq n} |x_k|$$

in (P), which is also known as *basis pursuit* and has been investigated widely in [11, 4, 10, 8, 9]. However, for  $0 < p < 1$  the  $\ell^p$ -quasi norm

$$\|x\|_p = \left( \sum_{k \leq n} |x_k|^p \right)^{1/p}, \quad 0 < p < 1$$

is closer to the  $\ell^0$ -function  $\|\cdot\|_0$  than  $\|\cdot\|_1$  is, see Figure 1. Therefore it is of natural interest to study problem (P) for  $f(x) = \|x\|_p$  which has also been done, for instance, in [13, 45, 40].

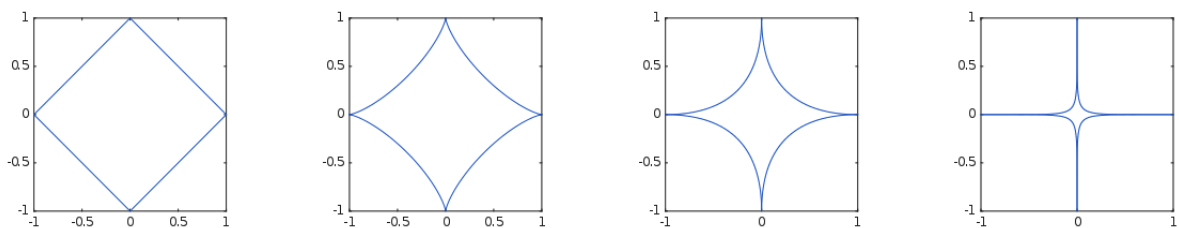


Figure 1: From  $\ell^1$  to  $\ell^0$ : Unit balls with respect to the  $\ell^p$ -quasi norm for decreasing  $p$ .

The cases (P) and  $(\text{P}^\varepsilon)$  with  $f(x) = \|x\|_p$  and  $0 \leq p \leq 1$  are usually called the *synthesis formulation* and expect the vector  $x$  to be sparse. However, it is observed in many applications, that the vector  $x$  is only *transform sparse*, i.e. sparse after an application of a suitable transform or often even only *compressible*, i.e. very few entries are large and the rest is small in modulus but not necessarily zero. For example, most natural images are *compressible* with respect to multiscale systems such as wavelets [16], curvelets [5] or shearlets [31]. More precisely, the transform coefficients decay to zero with a certain rate as the scales increase, see Figure 2. The transform sparse model is for instance used in *magnetic resonance imaging* (MRI) where the sparsity of medical images with respect to a wavelet transform is utilized cf. [39].

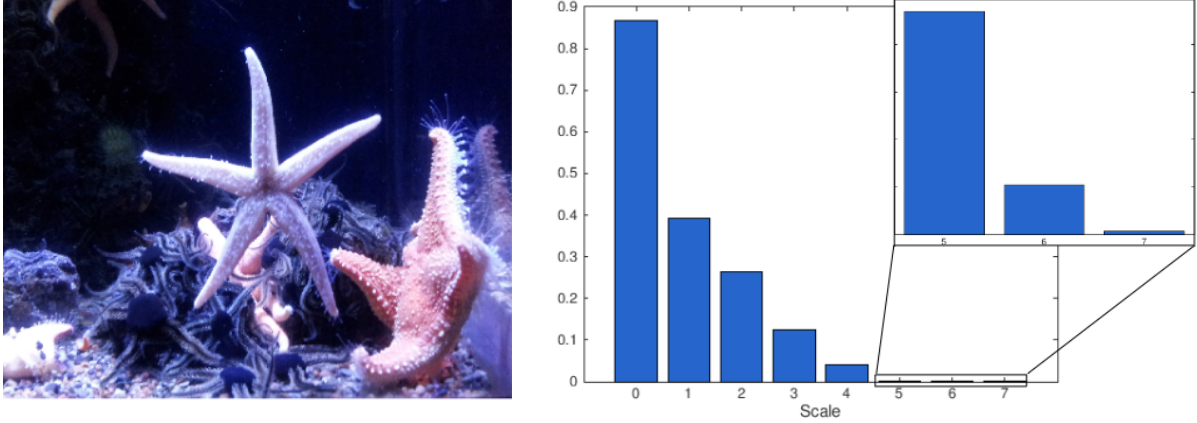


Figure 2: **Left:** A natural image of size  $2048 \times 2048$ . **Right:** Distribution of the shearlet coefficients of the best  $N$ -term approximation using 5% of the largest coefficients in modulus.

The transform sparse model leads to the so-called *analysis formulation* that is problems (P) and  $(P^\varepsilon)$  with

$$f(x) = \|\Psi x\|_p = \left( \sum_{\lambda \leq N} |\langle x, \psi_\lambda \rangle|^p \right)^{1/p}, \quad 0 < p \leq 1,$$

where  $\Psi : \mathbb{R}^n \rightarrow \mathbb{R}^N, x \mapsto (\langle x, \psi_\lambda \rangle)_{\lambda \leq N}$  is the *analysis operator* of a spanning system  $(\psi_\lambda)_{\lambda \leq N}$  for  $\mathbb{R}^n$ .

### 1.1 Overview of related work

Problems of the form (P) has been studied a lot in the literature for many different type of functions  $f$ . We will now give an overview of those works that are most relevant and related to our work. We start with one of the earliest models in compressed sensing, that is the *synthesis approach*.

#### Synthesis approach for $\ell^1$

The synthesis formulation for the  $\ell^1$ -minimization problem is  $(P^\varepsilon)$  with  $f(\cdot) = \|\cdot\|_1$ , i.e.

$$\min_x \|x\|_1 \quad \text{subject to} \quad \|y - Ax\|_2 \leq \varepsilon. \quad (\ell^1\text{-}P_S^\varepsilon)$$

The minimization problem  $(\ell^1\text{-}P_S^\varepsilon)$  is also known as (*inequality constrained*) *basis pursuit* and is very well studied in compressed sensing. Stability results of the form

$$\|x^* - x\|_2 \leq C_1 \varepsilon + C_2 \frac{\|x - x_s\|_1}{\sqrt{s}} \quad (1.2)$$

where  $x^*$  denotes the solution to  $(\ell^1\text{-}P_S^\varepsilon)$  and  $x_s$  is the best  $s$ -term approximation, i.e. the vector consisting of the  $s$  largest entries of  $x$ , are known. This is for example the case, if  $A$  satisfies the so-called *restricted isometry property* (RIP) [8, 20, 22]. We recall its definition here for convenience.

**Definition 1.1.** Let  $A \in \mathbb{C}^{m \times n}$  be a measurement matrix. If there exists  $\delta_s$  such that

$$(1 - \delta_s) \|x\|_2^2 \leq \|Ax\|_2^2 \leq (1 + \delta_s) \|x\|_2^2, \quad (1.3)$$

for all  $s$ -sparse vectors  $x$ , then we say  $A$  satisfies the RIP of order  $s$  with RIP constant  $\delta_s$ .

Note that there are other properties used in the literature such as the *null space property* [15] (and altered variants) to prove different stability estimates, see, for instance, [22].

### Synthesis approach for $\ell^p$

It has been noticed that the  $\ell^p$ -quasi norms yield stronger emphasis on the sparsity and thus in [13, 45, 21] the authors have studied the  $\ell^p$ -minimization problem

$$\min_x \|x\|_p^p \quad \text{subject to} \quad \|y - Ax\|_2 \leq \varepsilon. \quad (\ell^p\text{-P}_S^\varepsilon)$$

Again, the results that are fundamental for the user of this minimization problem are stability results of the form

$$\|x - x^*\|_2^p \leq C_1(p)\varepsilon^p + C_2(p) \frac{\|x - x_s\|_p^p}{s^{1-p/2}}, \quad (1.4)$$

where we have used the notations from above. A significant difference to the bound obtained in (1.3) is that the constants  $C_1(p)$  and  $C_2(p)$  in (1.4) now depend on  $p \in (0, 1)$ . However, for  $p = 1$  the constants in both cases agree [13, 45, 21]. The assumption on the measurement matrix  $A$  so that (1.4) holds is again based on the restricted isometry property.

The reader might wonder if such results are of any use in practice as the  $\ell^p$ -quasi norms turn the problem into a non-convex minimization problem which are in general NP-hard [23], again. However, in the same work [23] the authors showed that local minima can be computed in polynomial time by using an interior point method. Further, its improvement over  $\ell^1$  has also been numerically demonstrated in that paper.

### Analysis approach for $\ell^1$

Both minimization problems ( $\ell^1\text{-P}_S^\varepsilon$ ) and ( $\ell^p\text{-P}_S^\varepsilon$ ) assume the signal  $x$  to be directly sparse, however, as we have already mentioned in the introduction this is often not the case in many practical problems. One possible adaption of the synthesis problem is the analysis formulation of the minimization problem ( $\ell^1\text{-P}_A^\varepsilon$ ), that is

$$\min_x \|\Psi x\|_1 \quad \text{subject to} \quad \|y - Ax\|_2 \leq \varepsilon, \quad (\ell^1\text{-P}_A^\varepsilon)$$

where  $\Psi$  is some sparsifying transform associated to a spanning system  $(\psi_\lambda)_{\lambda \leq N}$  of  $\mathbb{R}^n$ . This problem has initially been studied in [6] and the authors of that work proved stability results of the form

$$\|x^* - x\|_2 \leq C_1\varepsilon + C_2 \frac{\|\Psi x - (\Psi x)_s\|_1}{\sqrt{s}} \quad (1.5)$$

for the case that  $\Psi$  is the *analysis operator* associated to a *Parseval frame* and the measurement matrix  $A$  satisfies the so-called  $\Psi$ -RIP.

**Definition 1.2.** Let  $\Psi \in \mathbb{C}^{N \times n}$  be a dictionary and  $A \in \mathbb{C}^{m \times n}$  the measurement matrix. If there exists  $\delta_s$  such that

$$(1 - \delta_s)\|\Psi^* x\|_2^2 \leq \|A\Psi^* x\|_2^2 \leq (1 + \delta_s)\|\Psi^* x\|_2^2, \quad (1.6)$$

for all  $s$ -sparse vectors  $x$ , then we say  $A$  satisfies the  $\Psi$ -RIP of order  $s$  with  $\Psi$ -RIP constant  $\delta_s$ .

This definition of the  $\Psi$ -RIP is the canonical analog to the restricted isometry property for direct sparse signals. Since its invention an avalanche of research took place producing a wide and fruitful literature on this topic. For instance, the relationship between the analysis formulation and the synthesis formulation has been studied in [18, 42] as well as possible generalizations of the results of [6] to general frames [28, 20, 27, 1, 47] and not only the case where  $\Psi$  corresponds to a Parseval frame. Such generalizations are of great importance as

the they show that not all frames are equally well suited for compressed sensing. By that we mean the number of measurements might vary significantly depending on the frame bounds, which is clearly not an effect that enters the argument if one is dealing with Parseval frames. This observation already hints that the choice of the sparsifying transform is a very delicate problem. We will next discuss what we are interested in in this work as well as what has to be carefully considered in some practical problems in that relation.

### Sparsifying transforms associated to arbitrary redundant systems

In the above stated stability estimates (1.2), (1.4) and (1.5) it is desirable to have a fast decay of  $\|x - x_s\|_1$ ,  $\|x - x_s\|_p^p$  or  $\|\Psi x - (\Psi x)_s\|_1$  respectively. When focusing on the latter expression it is implied that the transform coefficients  $(\langle x, \psi_\lambda \rangle)_\lambda$  should decrease fast in magnitude in order to have a meaningful use of such a stability estimate. Recall that such an estimate was obtained using the  $\Psi$ -RIP. Therefore the desire of having a fast decay of the transform coefficients is needed on the one hand, however, on the other hand we also want to have the  $\Psi$ -RIP to be fulfilled. The attentive reader will have noticed that there is a mismatch in this assumption. In fact, any signal of a given Hilbert space  $\mathcal{H}$  can be written as

$$x = \sum_{\lambda} \langle x, \psi_\lambda \rangle \tilde{\psi}_\lambda = \sum_{\lambda} \langle x, \tilde{\psi}_\lambda \rangle \psi_\lambda, \quad (1.7)$$

where  $(\psi_\lambda)_\lambda$  is a frame for the Hilbert space  $\mathcal{H}$  and  $(\tilde{\psi}_\lambda)_\lambda$  is a *dual frame*, see [14]. In particular, we have

$$x \neq \sum_{\lambda} \langle x, \psi_\lambda \rangle \psi_\lambda \quad (1.8)$$

in general for non-tight frames. Equations (1.7) and (1.8) show that if we require a fast decay of the coefficients  $(\Psi x)_s = (\langle x, \psi_\lambda \rangle)_s$ , then we must not assume the  $\Psi$ -RIP but rather the  $\tilde{\Psi}$ -RIP which is the  $\Psi$ -RIP with respect to the dual frame  $(\tilde{\psi}_\lambda)_\lambda$  and not the primal frame  $(\psi_\lambda)_\lambda$ . This rather simple observation can yield a great problem in some cases. For example, it is not unusual that although a primal frame is known explicitly as well as a certain behaviour of the transform coefficients  $(\langle x, \psi_\lambda \rangle)_\lambda$  but a dual  $(\tilde{\psi}_\lambda)_\lambda$  is not explicitly given. Moreover, the theoretical behaviour of the dual coefficients can be completely different to the primal frame coefficients. The shearlet transform [26, 35, 36, 37] is for instance a sparsifying transform that is often used in certain imaging applications [41, 44] as a sparsifying transform but an explicit formular of a dual does not exist – for the case of compactly supported shearlets, as the band-limited ones form a Parseval frame.

This issue makes an assumption such as the restricted isometry property of the dual frame rather artificial. Switching the roles, this means assuming the  $\Psi$ -RIP for the primal to hold, would on the other hand necessarily yield the minimization over dual frame coefficients. This again raises the question how these coefficients behave in general and if a dual is not known, then there is no point in minimizing over dual frame coefficients.

Finally, we want to remark that if the computation of a dual is feasible, one could also minimize over all duals in order to obtain the sparsest one. That again is the same as doing the synthesis formulation [38]. However, we are interested in the analysis formulation in this paper and a discussion of the analysis versus synthesis formulation is not in the focus of this paper, for further interested on this matter we refer the reader to [42].

## 1.2 Contribution

In this paper we combine  $(\ell^1\text{-P}_A^\varepsilon)$  and  $(\ell^p\text{-P}_S^\varepsilon)$  in order to obtain the best of both methods. While aiming for stability results we will also resolve the sparsity mismatch explained in the

previous section that arises when arbitrary frames are considered. Note that there are two fundamentally different possibilities to combine the two problems  $(\ell^1\text{-P}_\text{A}^\varepsilon)$  and  $(\ell^p\text{-P}_\text{S}^\varepsilon)$ . Either one considers

$$\min_x \|\Psi x\|_p^p \quad \text{subject to} \quad \|y - Ax\|_2 \leq \varepsilon \quad (\ell^p\text{-P}_\text{A}^\varepsilon)$$

with a transform  $\Psi$  that is associated to a fixed redundant system  $(\psi_\lambda)_\lambda$  or

$$\min_x \|\tilde{\Psi} x\|_p^p \quad \text{subject to} \quad \|y - Ax\|_2 \leq \varepsilon \quad (\ell^p\text{-}\widetilde{\text{P}}_\text{A}^\varepsilon)$$

which corresponds to the minimization over dual frame coefficients. The latter problem suggests to assume the  $\Psi$ -RIP with respect to the primal frame as then the sparsity pattern matches, cf. (1.7). Although stability is then very much expected we shall show that this is indeed the case, for the sake of completeness. For  $(\ell^p\text{-P}_\text{A}^\varepsilon)$  the situation is different, although it comes without any surprise that a stability can again be obtained if one assume the  $\Psi$ -RIP with respect to the dual frame, but, we shall not do this. More importantly, we show that for certain types of frames the assumption of the  $\Psi$ -RIP with respect to the primal frame is enough even if one minimizes over the primal frame coefficients. The property that enables this approach are *frames that have an identifiable dual*, see Section 2.1.

Our results can be connected to the literature as follows:

- For  $p = 1$  our result generalizes the findings of [6] to the case of arbitrary redundant systems that are not necessarily a Parseval frame. However, if the system is a Parseval frame both results (and constants) agree.
- If  $\Psi$  comes from a Parseval frame, then, due to the fact that all constants are given explicitly, we can compare  $\ell^p$ -minimization with  $\ell^1$ -minimization. Therefore, by carefully handling the parameters we can obtain smaller constants as for  $\ell^1$ -minimization studied in [6].
- Frames that have an identifiable dual are, coincidentally, a generalization of *scalable frames* [34] which have arisen in the literature from a completely different perspective.

Further, in practice it has been observed that the sparsity assumption might be impracticable, meaning that the sparsity  $s$  of the transform coefficients is larger than the dimension of the object  $x$  that is to be reconstructed, if the transform is truly redundant. We present a numerical experiment that shows that this is rather an issue coming from the implementation of such transforms than the theory. This experiment in turn implies that there might a large gap between the theory and the applications of compressed sensing for redundant dictionaries, cf. Section 5.

### 1.3 Outline

In Section 2 we present the new principles that we developed in order to obtain stability results for the analysis formulation of  $\ell^p$ -minimization  $(\ell^p\text{-P}_\text{A}^\varepsilon)$  and the minimization over dual coefficients  $(\ell^p\text{-}\widetilde{\text{P}}_\text{A}^\varepsilon)$ . The proofs of our main results are presented in Section 3 and Section 4 consists of a numerical experiment simulating the applicability and possible benefits of considering  $\ell^p$ -minimization in practice. In Section 5 we discuss the role of redundancy and how it fits to the model of sparsity in compressed sensing for the case of the shearlet transform.

## Notation

Throughout this paper  $\Psi := (\psi_\lambda)_{\lambda \leq N} \subset \mathcal{H}$  denotes a *frame* for some Hilbert space  $\mathcal{H}$ , i.e. there exist (positive and finite) constants  $c_1$  and  $c_2$  such that

$$c_1 \|x\|^2 \leq \sum_{i=1}^N |\langle x, \psi_\lambda \rangle|^2 \leq c_2 \|x\|^2 \quad \forall x \in \mathcal{H}.$$

The constants  $c_1$  and  $c_2$  are called *lower frame bound* and *upper frame bound*, respectively. If  $c_1 = c_2 = 1$ , then  $\Psi$  is called *Parseval frame*. With a slight abuse of notation we also denote the *analysis operator* by

$$\Psi : \mathcal{H} \rightarrow \mathbb{C}^N, \quad x \mapsto (\langle x, \psi_\lambda \rangle)_{\lambda \leq N},$$

and the synthesis operator  $\Psi^*$  by

$$\Psi^* : \mathbb{C}^N \rightarrow \mathcal{H}, \quad (c_\lambda)_{\lambda \leq N} \mapsto \sum_{\lambda \leq N} c_\lambda \psi_\lambda.$$

In most cases we will have  $\mathcal{H} = \mathbb{C}^n$ . If  $x$  is supposed to be an image in  $\mathbb{C}^{n \times n}$  we will consider  $\mathcal{H} = \mathbb{C}^{n^2}$  the vectorization of the  $n$  by  $n$  images.

Note that if the elements  $\psi_1, \dots, \psi_N$  are given as vectors in  $\mathbb{C}^{n^2}$ , then the synthesis operator is simply the matrix  $\Psi^* \in \mathbb{C}^{n^2 \times N}$  with columns  $\psi_\lambda, \lambda \leq N$ .

## 2 $\ell^p$ -minimization: the analysis formulation

This section contains the main concepts and results of this paper. We proceed by first proposing a new concept that is used later on to resolve the sparsity mismatch.

### 2.1 Identifiable duals

Let us recall the issue mentioned in the introduction. In the minimization problem ( $\ell^p$ -P<sub>A</sub><sup>ε</sup>) we are minimizing over primal frame coefficients  $(c_\lambda)_{\lambda \leq N} = (\langle x, \psi_\lambda \rangle)_{\lambda \leq N}$  and suppose we wish to prove stability estimates based on the  $\Psi$ -RIP. Then we notice that the  $\Psi$ -RIP is based on the sparsity of signals in the reconstruction system  $\Psi = (\psi_\lambda)_{\lambda \leq N}$  due to the synthesis operation which is the adjoint but not an inverse in the general frame case. More precisely, the following problem arises: The (assumed) sparse coefficient vector  $(c_\lambda)_\lambda$  that we obtain from the minimization problem does not give rise to a sparse representation with respect to the reconstruction system  $\Psi$  but rather to a dual  $\tilde{\Psi}$ , cf. (1.7) and (1.8). In fact, the dual frame coefficients might have a completely different sparsity pattern. In order to circumvent this problem we make the following definition.

**Definition 2.1.** *Let  $I$  be some index set. We say a frame  $\Psi = (\psi_\lambda)_{\lambda \in I}$  for  $\mathcal{H}$  has an identifiable dual if there exists a dual frame  $\tilde{\Psi}$  such that for all  $x \in \mathcal{H}$  and any  $\lambda \in I$  the coefficient in modulus  $|\langle x, \tilde{\psi}_\lambda \rangle|$  can be bounded from above and below by  $|\langle x, \psi_\lambda \rangle|$ , i.e. there exist constants  $d_1, d_2 > 0$  such that*

$$d_1 |\langle x, \psi_\lambda \rangle| \leq |\langle x, \tilde{\psi}_\lambda \rangle| \leq d_2 |\langle x, \psi_\lambda \rangle| \quad (2.1)$$

for all  $x \in \mathcal{H}$  and all  $\lambda \in I$ .

If  $\Psi$  is a frame that has an identifiable dual, then this property ensures that there exists a dual such that the sparsity of the primal frame coefficients leads to a sparse representation in a dual system. We proceed with discussing its relation to other special frames.

It is obvious that every tight frame has an identifiable dual. More generally, every *scalable frame* (cf. Definition 2.2) has an identifiable dual.

**Definition 2.2** ([34]). A frame  $\Psi = \{\psi_\lambda : \lambda \in \mathbb{N}\}$  for  $\mathcal{H}$  is called *scalable* if there exists scalars  $c_\lambda \geq 0, \lambda \in \mathbb{N}$  such that

$$\{c_\lambda \psi_\lambda : \lambda \in \mathbb{N}\}$$

forms a Parseval frame for  $\mathcal{H}$ . If there exists  $\delta > 0$  such that  $c_j > 0$ , then we call  $\Psi$  *positively scalable*.

The following result is trivial to prove and we therefore omit its proof.

**Proposition 2.3.** *Every scalable frame has an identifiable dual.*

The question that naturally arises is whether the two concepts agree. This is indeed not the case as the following example shows.

**Example 2.4.** It is easy to check that the set of vectors

$$\Psi = \left\{ \begin{pmatrix} 2 \\ 1 \end{pmatrix}, \begin{pmatrix} 2 \\ 2 \end{pmatrix}, \begin{pmatrix} 1 \\ 2 \end{pmatrix} \right\}$$

is frame for  $\mathbb{R}^2$  that has an identifiable dual, e.g.

$$\tilde{\Psi} = \left\{ \begin{pmatrix} 2 \\ 1 \end{pmatrix}, \begin{pmatrix} -2 \\ -2 \end{pmatrix}, \begin{pmatrix} 1 \\ 2 \end{pmatrix} \right\}$$

with  $d_1 = d_2 = 1$  but it is neither a tight nor a scalable frame. The fact that it is not scalable follows from Theorem 3.6 of [34].

As we will also see later in this article the identifiability is actually not needed for the entire space but only to certain elements, cf. Remark 2.9

Next, we show how this concept relates to the *restricted isometry property* and in particular the number of measurements that are needed in order for it to hold.

## 2.2 Restricted isometry property

The restricted isometry property was first introduced by Candès and Tao in [10] based on the assumption that often the signal that is to be reconstructed is sparse. It was then later generalized by Candès et al. in [6] to transform sparse signals. This covers a large class of signals considered in imaging tasks, since the images considered are often sparse in, for instance, a wavelet domain.

**Definition 2.5.** Let  $\Psi \in \mathbb{C}^{N \times n}$  be a dictionary and  $A \in \mathbb{C}^{m \times n}$  the measurement matrix. If there exists  $\delta_s$  such that

$$(1 - \delta_s) \|\Psi^* x\|_2^2 \leq \|A \Psi^* x\|_2^2 \leq (1 + \delta_s) \|\Psi^* x\|_2^2, \quad (2.2)$$

for all  $s$ -sparse vectors  $x$ , then we say  $A$  satisfies the  $\Psi$ -RIP of order  $s$  with  $\Psi$ -RIP constant  $\delta_s$ .

The existence of matrices satisfying the  $\Psi$ -RIP is given by the following theorem which will also be needed for our main result. We move the proof to the appendix as it very close to the proof given in [30].

**Theorem 2.6.** Fix a probability measure  $\nu$  on  $\{1, \dots, N\}$ , a sparsity level  $s < N$ , and a constant  $0 < \delta < 1$ . Let  $\Psi = (\psi_\lambda)_{\lambda \leq N}$  be a frame with frame bounds  $c_1$  and  $c_2$  and let  $A$  be an  $n \times n$  matrix whose rows  $(r_i)_{i \leq n}$  satisfy

$$\sum_i r_k(i) r_j(i) \nu_i = \delta_{j,k}.$$



Furthermore, let  $K$  be a number such that  $\|\psi_\lambda\| \leq K$  for all  $\lambda \leq N$  and

$$L = \sup_{\substack{\|\Psi^* c\|=1 \\ \|c\|_0 \leq s}} \frac{\|(\Psi \Psi^* c)_\lambda\|_1}{\sqrt{s}}. \quad (2.3)$$

If  $\tilde{A}$  is an  $m \times n$  submatrix whose rows are subsampled from  $A$  according to  $\nu$ , then there exists a  $C > 0$  independent of all relevant parameters such that for

$$m \geq \frac{CK}{c_1} \delta^{-2} s L^2 \max\{\log^3(sL^2) \log(N), \log(1/\gamma)\} \quad (2.4)$$

the normalized submatrix  $\sqrt{\frac{1}{m}} \tilde{A}$  satisfies the  $\Psi$ -RIP of order  $s$  with  $\Psi$ -RIP constant  $\delta$  with probability  $1 - \gamma$ . Furthermore, if the frame possess an identifiable dual with upper constant  $d_2$ , then  $1/c_1$  can be replaced by  $d_2$  in (2.4), i.e. the result holds for

$$m \geq d_2 CK \delta^{-2} s L^2 \max\{\log^3(sL^2) \log(N), \log(1/\gamma)\} \quad (2.5)$$

Theorem 2.6 states that the  $\Psi$ -RIP can be guaranteed provided the number of measurements scales properly with the sparsity  $s$ . In fact, there are some further subtleties that have to be considered. First, it is not only the sparsity  $s$  that is important but also the localization factor  $L$ . This factor can be controlled, for instance, if the frame is *localized*, [19]. In particular, if the Gramian has a strong off-diagonal decay, then  $L$  can be further characterized with respect to this decay rate as  $L$  measures how the Gramian distorts the sparsity structure. Second, the dependency of the lower frame bound by  $1/c_1$  can also be controlled by localization arguments. Note that by Greshgorin's Circle Theorem there exists a  $\lambda \in \{1, \dots, N\}$  such that

$$\left| \frac{1}{c_1} - \|\psi_\lambda\|^2 \right| \leq \sum_{\mu \neq \lambda} |\langle \tilde{\psi}_\lambda, \tilde{\psi}_\mu \rangle|$$

and therefore

$$\frac{1}{c_1} \leq \sum_{\mu \leq N} |\langle \tilde{\psi}_\lambda, \tilde{\psi}_\mu \rangle|.$$

Thus frames with strong localization properties such that the dual frame is also strongly localized are of particular interest. It was also proven in [19] that if the primal frame  $(\psi_\lambda)_{\lambda \in \mathbb{N}}$  is *polynomially self-localized*, i.e. there exists  $C > 0$  such that

$$|\langle \psi_\lambda, \psi_\mu \rangle| \leq C \frac{1}{(1 + |\lambda - \mu|)^r},$$

for some  $r \in \mathbb{N}$  and all  $\lambda \neq \mu$ , then the dual frame  $(\tilde{\psi}_\lambda)_{\lambda \in \mathbb{N}}$  will also be polynomially self-localized. This result was extended in [24] to more general geometries, i.e. to other index functions than the euclidean distance.

Theoretically,  $s$  can be considered as the dominating value for the number of measurements. However, in practice it is a priori not clear that it can always be assumed to be smaller than  $n$ . In fact, since  $N$  is usually much much larger than  $n$  the analysis coefficients must be extremely sparse to make this result useful. In Section 5 we will demonstrate that this is indeed an issue in many imaging problems, but we also show that it has to be further studied from a practical point of view and is not really an issue of the theory.

The last concept that we need before we can prove the stability result is the notion of *stable frame bounds*. This is merely a concept used in order to make all constants and variables feasible. It also shows how large  $s$  can be in the worst case.

### 2.3 Stable frame bounds

As we are now not necessarily dealing with tight frames or Parseval frames, the ratio of the frame bounds  $c_1/c_2$  is not constant and in fact plays a vital role in the estimates for the proof of our main results. In order to have some control over this ratio we make the following definition.

**Definition 2.7.** *We say a frame has  $q$ -controllable frame bounds if there exists a  $q \in \{1, \dots, N\}$  such that*

$$\frac{c_1^{p/(2-p)}}{c_2^{p/(2-p)}} \geq \frac{q}{N}. \quad (2.6)$$

Note that (2.6) is, same as the identifiability condition in Section 2.1, always fulfilled for tight frames, in particular it holds for every orthonormal basis or Parseval frame.

We will need the controllability of the frame bound in order to make the implicit assumptions on the range of the sparsity  $s$  more precise, cf. Theorem 2.8.

Furthermore, the controllability of the frame bounds is also of numerical importance since the number of frame elements  $N$  is usually very large in practice and in order to have numerical stability the frame ratio  $c_1/c_2$  should not be too small. In Section 4 we will verify (2.6) for *digital shearlets*, [31]. Note that the term on the left-hand side of 2.6 decreases as  $p$  increases.

### 2.4 Stability for analysis formulation

Based on the previously introduced principles of identifiable duals and controllable frame bounds, we can prove the following theorem that guarantees stability of solutions obtained by  $(\ell^p\text{-P}_A^\varepsilon)$ . The proof of the theorem is postponed to Section 3.

**Theorem 2.8.** *Let  $\Psi$  be a frame that has  $q$ -controllable frame bounds and has an identifiable dual. Moreover, let  $A$  be a measurement matrix satisfying the  $\Psi$ -RIP with  $\delta_\nu < 0.5$  where  $\nu = s \frac{2 \cdot (1+2^{-p})^{1/(p-2)} c_2^{p/(p-2)} d_2^{p/(p-2)}}{c_1^{p/(p-2)}}$  and  $s < \frac{q}{2d_2^2(1+2^{-p})^{1/(p-2)}}$  and  $d_2 > 0$  is the upper constant in the identifiability conditions. Then the solution  $x^*$  of  $(\ell^p\text{-P}_A^\varepsilon)$  satisfies*

$$\|x - x^*\|_2^p \leq C_1(p)\varepsilon^p + C_2(p) \frac{\|\Psi x - (\Psi x)_s\|_p^p}{s^{1-p/2}}$$

for some positive constants  $C_1(p)$  and  $C_2(p)$  that depend on  $p$ , the frame bounds  $c_1, c_2$ , the  $\Psi$ -constant  $\delta_\nu$ , the sparsity  $s$ , the controllability parameter  $q$ , and the constants from the identifiability condition.

**Remark 2.9.** *The proof of Theorem 2.6 and Theorem 2.8 show that we use the identifiability only for the rows of the measurement matrix and  $x - x^*$ . Therefore, it does not need to hold for every element of  $\mathbb{C}^n$ . Further, the dual does not have to be known or constructed explicitly. It suffices to know that it exists.*

For  $p = 1$  Theorem 2.8 generalizes the findings of [6] to the case of non-tight redundant systems, thus our result can be seen as a generalization of the stability result for the analysis formulation of  $\ell^1$ -minimization ( $\ell^1\text{-P}_A^\varepsilon$ ). In particular, if  $\Psi$  is a Parseval frame the constants  $C_1(p), C_2(p)$  agree for  $p = 1$  with those obtained in [6]. Furthermore, the constants  $C_1(p), C_2(p)$  appearing in the proof are monotonically decreasing for decreasing  $p$ . We state this fact as a corollary.

**Corollary 2.10.** *Let  $\Psi$  be a Parseval frame and let  $A$  be a measurement matrix satisfying the  $\Psi$ -RIP with  $\delta_{7s} < 0.6$  and  $s < \frac{q}{2 \cdot (1+2^{-p})^{1/(p-2)}}$ . Then the solution  $x^*$  of  $(\ell^p\text{-P}_A^\varepsilon)$  satisfies*

$$\|x - x^*\|_2^p \leq C_1(p)\varepsilon^p + C_2(p) \frac{\|\Psi x - (\Psi x)_s\|_p^p}{s^{1-p/2}}$$

for some positive constants  $C_1(p)$  and  $C_2(p)$  that depend on  $p$ . Moreover, the constants can be chosen to be monotonically decreasing for decreasing  $p$  with maximal constants agreeing with those obtained in [6] for  $p = 1$ .

For the sake of completeness we also discuss the minimization problem  $(\ell^p\text{-}\widetilde{\mathbf{P}}_A^\varepsilon)$  while assuming the  $\Psi$ -RIP.

## 2.5 Stability for dual analysis formulation

As we already mentioned in Section 2.1, our motivation for the concept of an identifiable dual was to deal with the mismatch between the sparsity system  $(\psi_\lambda)_{\lambda \leq N}$  and the coefficients  $(\langle x, \psi_\lambda \rangle)_{\lambda \leq N}$  which do not give rise to a representation in  $(\psi_\lambda)_{\lambda \leq N}$ . It is therefore very natural to expect when minimizing over the dual frame coefficients, the concept of an identifiable dual is not needed. This is indeed the case as the following theorem shows.

**Theorem 2.11.** *Let  $A$  be a measurement matrix satisfying the  $\Psi$ -RIP with  $\delta_\nu < 0.5$  where  $\nu = s \left( \frac{3}{c_1^{2p}} \right)^{1/(2-p)}$  and  $s < N \left( \frac{c_1^{2p}}{3} \right)^{1/(2-p)}$ . Then the solution  $x^*$  of  $(\ell^p\text{-}\widetilde{\mathbf{P}}_A^\varepsilon)$  satisfies*

$$\|x - x^*\|_2^p \leq C_1(p)\varepsilon^p + C_2(p) \frac{\|\widetilde{\Psi}x - (\widetilde{\Psi}x)_s\|_p^p}{s^{1-p/2}}$$

for some positive constants  $C_1(p)$  and  $C_2(p)$  that depend on  $p$ , the frame bounds  $c_1, c_2$ , the  $\Psi$ -RIP constant  $\delta_\nu$ , and the sparsity  $s$ .

The reader might wonder why the controllability of the frame bounds is not assumed. In fact, the role of the controllability parameter  $q$  was to determine the range of  $s$ . However, loosely speaking, the constants are now appearing in the right order which makes the assumptions superfluous and the range of  $s$  can be determined entirely.

The proof follows very closely the lines of the proof of Theorem 2.8. We provide a sketch of it in Subsection 3.3.

## 3 Proofs

### 3.1 Proof of Theorem 2.8

The proof of Theorem 2.8 is mainly inspired by [6] and [45] which have their roots in [8].

Let  $x, x^*$  be as in the theorem and define  $z = x - x^*$ . Further, set  $T_0$  as the set consisting of the  $s$  largest coefficients of  $\Psi x$  in  $p$ -th modulus. For any set  $T$ ,  $\Psi_T$  should denote the matrix restricted to columns indexed by  $T$ . Then we divide  $T_0^c$  into sets  $T_1, T_2, \dots$  of size  $M$  in order of decreasing magnitude of  $\Psi_{T_0^c} z$ . The explicit value of  $M$  will be determined later. We shall need the following results.

The first one, Lemma 3.1 below, is a trivial modification of the *cone constraint* Lemma 2.1 in [6]. For completeness we provide a proof.

**Lemma 3.1.** *The vector  $\Psi z$  obeys the following cone constrained*

$$\|\Psi_{T_0^c} z\|_p^p \leq 2\|\Psi_{T_0^c} x\|_p^p + \|\Psi_{T_0} z\|_p^p.$$

*Proof.* Since  $x$  and  $x^*$  are both feasible and  $x^*$  is the minimizer of  $(\ell^p\text{-}\mathbf{P}_A^\varepsilon)$  we have

$$\begin{aligned} \|\Psi_{T_0} x\|_p^p + \|\Psi_{T_0^c} x\|_p^p &\geq \|\Psi x\|_p^p \\ &\geq \|\Psi x^*\|_p^p \\ &= \|\Psi x - \Psi z\|_p^p \end{aligned}$$

$$\geq \|\Psi_{T_0}x\|_p^p - \|\Psi_{T_0^c}x\|_p^p - \|\Psi_{T_0}z\|_p^p + \|\Psi_{T_0^c}z\|_p^p.$$

□

The following lemma generalizes Lemma 2.2 in [6] to the case  $p \in (0, 1)$ . For  $p = 1$  the results coincide.

**Lemma 3.2.** *For  $\rho = s/M$  and  $\eta = 2\|\Psi_{T_0^c}x\|_p^p/s^{1-p/2}$  we have*

$$\sum_{j \geq 2} \|\Psi_{T_j}z\|_2^p \leq \rho^{1-p/2}(\|\Psi_{T_0}z\|_2^p + \eta).$$

*Proof.* By construction we have that every entry in  $|(\Psi_{T_{j+1}}z)|^p$  which will be denoted by  $|(\Psi_{T_{j+1}}z)|_{(k)}^p$  is bounded as follows

$$|(\Psi_{T_{j+1}}z)|_{(k)}^p \leq \frac{\|\Psi_{T_j}z\|_p^p}{M}.$$

Therefore

$$\|\Psi_{T_{j+1}}z\|_2^2 = \sum_{k=1}^M |(\Psi_{T_{j+1}}z)|_{(k)}^2 \leq M^{1-2/p} \|\Psi_{T_j}z\|_p^2$$

and thus

$$\sum_{j \geq 2} \|\Psi_{T_j}z\|_2^p \leq \sum_{j \geq 1} \frac{\|\Psi_{T_j}z\|_p^p}{M^{1-p/2}} = \frac{\|\Psi_{T_0^c}z\|_p^p}{M^{1-p/2}}.$$

Applying Lemma 3.1 and the generalized mean inequality gives

$$\sum_{j \geq 2} \|\Psi_{T_j}z\|_2^p \leq \left(\frac{s}{M}\right)^{1-p/2} \left(\|\Psi_{T_0}z\|_2^p + \frac{2\|\Psi_{T_0^c}x\|_p^p}{s^{1-p/2}}\right).$$

□

**Lemma 3.3** ([6], Lemma 2.3). *The vector  $Az$  satisfies*

$$\|Az\|_2 \leq 2\varepsilon.$$

**Lemma 3.4.** *Let  $c_2$  be the upper frame bound for the frame  $\Psi$ . Then we have*

$$(1 - \delta_{s+M})^{p/2} \|(\Psi_{T_{01}})^* \Psi_{T_{01}} \tilde{z}\|_2^p - (c_2(1 + \delta_M))^{p/2} \rho^{1-p/2} \left( c_2^{p/2} d_2^p \|z\|_2^p + \eta \right) \leq (2\varepsilon)^p$$

where  $T_{01} := T_0 \cup T_1$ .

*Proof.* W.l.o.g. we assume the identifiable dual is the canonical dual and define  $\tilde{z} := (\Psi^* \Psi)^{-1} z$ , note that  $\Psi^* \Psi$  is invertible as it is the frame operator. Then by using Lemma 3.3 we have

$$(2\varepsilon)^p \geq \|Az\|_2^p \tag{3.1}$$

$$= \|A\Psi^* \Psi \tilde{z}\|_2^p \geq \|A(\Psi_{T_{01}})^* \Psi_{T_{01}} \tilde{z}\|_2^p \tag{3.2}$$

$$- \sum_{j \geq 2} \|A(\Psi_{T_j})^* \Psi_{T_j} \tilde{z}\|_2^p. \tag{3.3}$$

and by the  $\Psi$ -RIP we have

$$(1 - \delta_{s+M})^{p/2} \|(\Psi_{T_{01}})^* \Psi_{T_{01}} \tilde{z}\|_2^p \leq \|A(\Psi_{T_{01}})^* \Psi_{T_{01}} \tilde{z}\|_2^p \quad (3.4)$$

and

$$\sum_{j \geq 2} \|A(\Psi_{T_j})^* \Psi_{T_j} \tilde{z}\|_2^p \leq (1 + \delta_M)^{p/2} \sum_{j \geq 2} \|(\Psi_{T_j})^* \Psi_{T_j} \tilde{z}\|_2^p. \quad (3.5)$$

Recall that by the identifiability there exists  $d_1, d_2$  be such that

$$d_1 |\langle z, \psi_\lambda \rangle| \leq |\langle z, \tilde{\psi}_\lambda \rangle| \leq d_2 |\langle z, \psi_\lambda \rangle| \quad \forall \lambda \leq N. \quad (3.6)$$

Using the frame property, the identifiability and Lemma 3.2 yields

$$\sum_{j \geq 2} \|(\Psi_{T_j})^* \Psi_{T_j} \tilde{z}\|_2^p \leq \sum_{j \geq 2} c_2^{p/2} d_2^p \|\Psi_{T_j} z\|_2^p \quad (3.7)$$

$$\leq c_2^{p/2} d_2^p \rho^{1-p/2} (\|\Psi_{T_0} z\|_p^p + \eta) \quad (3.8)$$

$$\leq c_2^{p/2} d_2^p \rho^{1-p/2} (c_2^{p/2} \|z\|_p^p + \eta). \quad (3.9)$$

Combing (3.3), (3.4), (3.5) and (3.9) yields the claim.  $\square$

**Lemma 3.5.** *The following inequality is true*

$$\|z\|_2^{2p} \leq \frac{1}{c_1^p d_1^{2p}} \left( \frac{1}{c_1^{p/2}} \|z\|_2^p \|(\Psi_{T_{01}})^* \Psi_{T_{01}} \tilde{z}\|_2^p + \rho^{2-p} (c_2^{p/2} \|z\|_2^p + \eta)^2 \right),$$

*Proof.* With  $T_{01} := T_0 \cup T_1$  we have by using the frame property and the identifiability of a dual

$$\begin{aligned} \|h\|_2^{2p} &\leq \frac{1}{c_1^p} \|\Psi z\|_2^{2p} \\ &\leq \frac{1}{c_1^p d_1^{2p}} \left( \|\Psi_{T_{01}} \tilde{z}\|_2^{2p} + \|\Psi_{T_{01}^c} z\|_2^{2p} \right) \\ &= \frac{1}{c_1^p d_1^{2p}} \left( (\langle \tilde{z}, (\Psi_{T_{01}})^* \Psi_{T_{01}} \tilde{z} \rangle)^p + \|\Psi_{T_{01}^c} z\|_2^{2p} \right) \\ &\leq \frac{1}{c_1^p d_1^{2p}} \left( \|\tilde{z}\|_2^p \|(\Psi_{T_{01}})^* \Psi_{T_{01}} \tilde{z}\|_2^p + \|\Psi_{T_{01}^c} z\|_2^{2p} \right) \\ &\leq \frac{1}{c_1^p d_1^{2p}} \left( \frac{1}{c_1^{p/2}} \|z\|_2^p \|(\Psi_{T_{01}})^* \Psi_{T_{01}} \tilde{z}\|_2^p + \rho^{2-p} (\|\Psi_{T_0} z\|_2^p + \eta)^2 \right), \end{aligned}$$

where the last inequality follows from Lemma 3.2.  $\square$

The proof of the main result can now be derived from the previous lemmas.

*Proof of Theorem 2.8.* By Lemma 3.5 we have

$$\begin{aligned} \|z\|_2^{2p} &\leq \frac{1}{c_1^p d_1^{2p}} \left( \frac{1}{c_1^{p/2}} \left( \frac{\gamma_1 \|z\|_2^{2p}}{2} + \frac{\|(\Psi_{T_{01}})^* \Psi_{T_{01}} \tilde{z}\|_2^{2p}}{2\gamma_1} \right) + \rho^{2-p} (c_2^{p/2} \|z\|_2^p + \eta)^2 \right) \\ &= \frac{1}{c_1^p d_1^{2p}} \left( \frac{1}{c_1^{p/2}} \left( \frac{\gamma_1 \|z\|_2^{2p}}{2} + \frac{\|(\Psi_{T_{01}})^* \Psi_{T_{01}} \tilde{z}\|_2^{2p}}{2\gamma_1} \right) + \rho^{2-p} \left( c_2^{p/2} \|z\|_2^p + 2c_2^{p/2} d_2^p \|z\|_2^p \eta + \eta^2 \right) \right) \\ &\leq \frac{1}{c_1^p d_1^{2p}} \left( \frac{1}{c_1^{p/2}} \left( \frac{\gamma_1 \|z\|_2^{2p}}{2} + \frac{\|(\Psi_{T_{01}})^* \Psi_{T_{01}} \tilde{z}\|_2^{2p}}{2\gamma_1} \right) + \rho^{2-p} \left( c_2^{p/2} \|z\|_2^p + \left( c_2^{p/2} d_2^p \gamma_2 \|z\|_2^p + \frac{\eta^2}{\gamma_2} \right) + \eta^2 \right) \right) \end{aligned}$$

$$= \frac{1}{c_1^p d_1^{2p}} \left( \frac{1}{c_1^{p/2}} \left( \frac{\gamma_1 \|z\|^{2p}}{2} + \frac{\|(\Psi_{T_{01}})^* \Psi_{T_{01}} \tilde{z}\|^{2p}}{2\gamma_1} \right) + \rho^{2-p} \left( c_2^p d_2^{2p} (1 + \gamma_2) \|z\|_2^{2p} + \left( \frac{1}{\gamma_2} + 1 \right) \eta^2 \right) \right).$$

Therefore

$$\left( 1 - \frac{1}{c_1^p d_1^{2p}} \left( \frac{\gamma_1}{2c_1^{p/2}} + \rho^{2-p} c_2^p d_2^{2p} (1 + \gamma_2) \right) \right) \|z\|_2^{2p} \leq \frac{1}{c_1^p d_1^{2p}} \left( \frac{\|(\Psi_{T_{01}})^* \Psi_{T_{01}} \tilde{z}\|^{2p}}{2\gamma_1 c_1^{p/2}} + \rho^{2-p} \left( \frac{1}{\gamma_2} + 1 \right) \eta^2 \right).$$

and in particular

$$\left( c_1^p d_1^{2p} - \left( \frac{\gamma_1}{2c_1^{p/2}} + \rho^{2-p} c_2^p d_2^{2p} (1 + \gamma_2) \right) \right)^{1/2} \|z\|_2^p \leq \frac{\|(\Psi_{T_{01}})^* \Psi_{T_{01}} \tilde{z}\|^p}{\sqrt{2\gamma_1 c_1^{p/2}}} + \rho^{1-p/2} \left( \frac{1}{\gamma_2} + 1 \right)^{1/2} \eta.$$

and hence, by using Lemma 3.4 we obtain

$$\Gamma_1(p) \|z\|_2^p - \Gamma_2(p) \eta \leq (2\varepsilon)^p$$

with

$$\Gamma_1(p) = \sqrt{2\gamma_1 c_1^{p/2} (1 - \delta_{s+M})^p (c_1^p d_1^{2p} - (\frac{\gamma_1}{2c_1^{p/2}} + \rho^{2-p} c_2^p d_2^{2p} (1 + \gamma_2)))} - \sqrt{(c_2 d_2^2 (1 + \delta_M))^p \rho^{2-p}}$$

and

$$\Gamma_2(p) = \sqrt{2\gamma_1 c_1^{p/2} (1 - \delta_{s+M})^p \rho^{2-p} \left( \frac{1}{\gamma_2} + 1 \right)} - \sqrt{(c_2 d_2^2 (1 + \delta_M))^p \rho^{2-p}}.$$

It is left to choose  $\gamma_1, \gamma_2$ , and  $\rho$  so that  $\Gamma_1(p)$  and  $\Gamma_2(p)$  are non-negative. W.l.o.g. we can assume  $c_1 \geq 1$ , otherwise a rescaling can be performed. We now choose  $\gamma_1 = c_1^{3p/2} d_1^{2p}$  and  $M = s \frac{2 \cdot (1+2^{-p})^{1/(p-2)} c_2^{p/(p-2)} d_2^{2p/(p-2)}}{c_1^{p/(p-2)}}$ . Then  $\Gamma_1(p)$  is larger than zero and for  $\gamma_2$  sufficiently small  $\Gamma_2(p)$  is also larger than zero. Furthermore, one can check

$$\Gamma'_1(p) < 0 \quad \text{and} \quad \Gamma'_2(p) > 0$$

holds for  $\gamma_2$  small enough. This completes the proof.  $\square$

### 3.2 Proof of Corollary 2.10

The Proof of Corollary 2.10 follows the same lines as the Proof of Theorem 2.8 and we skip the details. We only have to verify the statement for regarding the constants. For  $c_1 = c_2 = 1$  the above computations reduce to

$$\Gamma_1(p) \|z\|_2^p - \Gamma_2(p) \eta \leq (2\varepsilon)^p$$

with the constants

$$\begin{aligned} \Gamma_1(p) &= \sqrt{2\gamma_1 (1 - \delta_{s+M})^p \left( 1 - \left( \frac{\gamma_1}{2} + \rho^{2-p} (1 + \gamma_2) \right) \right)} - \sqrt{((1 + \delta_M))^p \rho^{2-p}} \\ \Gamma_2(p) &= \sqrt{2\gamma_1 (1 - \delta_{s+M})^p \rho^{2-p} \left( \frac{1}{\gamma_2} + 1 \right)} - \sqrt{((1 + \delta_M))^p \rho^{2-p}}. \end{aligned}$$

One possibility to guarantee  $\Gamma_1(p)$  and  $\Gamma_2(p)$  to be positive is to choose  $\gamma_1 = 1, M = 6s$  then  $\Gamma_1(p)$  is positive and for  $\gamma_2$  sufficiently small  $\Gamma_2(p)$  is positive. Moreover, note that

$$\Gamma'_1(p) < 0 \quad \text{and} \quad \Gamma'_2(p) > 0$$

for  $\gamma_2$  small enough.

### 3.3 Proof of Theorem 2.11

The proof follows the argumentation shown in Section 3.1. We will summarize the results that are needed and sketch a proof of the result.

**Proposition 3.6.** *Retaining the notations and definition Subsection 3.1 the following estimates hold*

i)

$$\|\tilde{\Psi}_{T_0^c} z\|_p^p \leq 2\|\tilde{\Psi}_{T_0^c} x\|_p^p + \|\tilde{\Psi}_{T_0} z\|_p^p.$$

ii)

$$\sum_{j \geq 2} \|\tilde{\Psi}_{T_j} z\|_2^p \leq \rho^{1-p/2} (\|\tilde{\Psi}_{T_0} z\|_2^p + \eta).$$

iii)

$$(1 - \delta_{s+M})^{p/2} \|(\Psi_{T_{01}})^* \Psi_{T_{01}} \tilde{z}\|_2^p - (c_2(1 + \delta_M))^{p/2} \rho^{1-p/2} \left( c_1^{-p/2} \|z\|_2^p + \eta \right) \leq (2\varepsilon)^p$$

iv)

$$\|z\|_2^{2p} \leq c_2^p \left( \|z\|_2^p \|(\tilde{\Psi}_{T_{01}})^* \tilde{\Psi}_{T_{01}} \tilde{z}\|_2^p + \rho^{2-p} (\|\tilde{\Psi}_{T_0} z\|_2^p + \eta)^2 \right).$$

*Proof.* All properties are proved in the same manner as in Subsection 3.1, in particular, item i), ii), and iv) are trivial adaptations and will be skipped. As for item iii), note that

$$\begin{aligned} (2\varepsilon)^p &\geq \|Az\|_2^p \\ &\geq \|A(\Psi_{T_{01}})^* \Psi_{T_{01}} \tilde{z}\|_2^p - \sum_{j \geq 2} \|A(\Psi_{T_j})^* \Psi_{T_j} \tilde{z}\|_2^p \\ &\geq (1 - \delta_{s+M})^{p/2} \|(\Psi_{T_{01}})^* \Psi_{T_{01}} \tilde{z}\|_2^p - (1 + \delta_M)^{p/2} \sum_{j \geq 2} \|(\Psi_{T_j})^* \Psi_{T_j} \tilde{z}\|_2^p. \end{aligned}$$

By ii) we have

$$\begin{aligned} \sum_{j \geq 2} \|(\Psi_{T_j})^* \Psi_{T_j} \tilde{z}\|_2^p &\leq c_2^{p/2} \sum_{j \geq 2} \|\Psi_{T_j} \tilde{z}\|_2^p \\ &= c_2^{p/2} \sum_{j \geq 2} \|\tilde{\Psi}_{T_j} z\|_2^p \\ &\leq c_2^{p/2} \rho^{1-p/2} (\|\tilde{\Psi}_{T_0} z\|_2^p + \eta). \end{aligned}$$

Thus the claim follows. □

*Sketch of Proof of Theorem 2.11.* By Proposition 3.6 iv) we have

$$\begin{aligned} \|z\|_2^{2p} &\leq c_2^p \left[ \left( \frac{\gamma_1 \|z\|_2^{2p}}{2} + \frac{\|(\tilde{\Psi}_{T_{01}})^* \tilde{\Psi}_{T_{01}} \tilde{z}\|_2^p}{2\gamma_1} \right) + \rho^{2-p} (\|\tilde{\Psi}_{T_{01}} z\|_2^{2p} + 2\|\tilde{\Psi}_{T_0} z\|_2^p \eta + \eta^2) \right] \\ &\leq c_2^p \left[ \left( \frac{\gamma_1 \|z\|_2^{2p}}{2} + \frac{\|(\tilde{\Psi}_{T_{01}})^* \tilde{\Psi}_{T_{01}} \tilde{z}\|_2^p}{2\gamma_1} \right) + \rho^{2-p} \left( \|\tilde{\Psi}_{T_{01}} z\|_2^{2p} + \eta^2 + \left( \gamma_2 \|\tilde{\Psi}_{T_0} z\|_2^{2p} + \frac{\eta^2}{\gamma_2} \right) \right) \right] \end{aligned}$$

$$\leq c_2^p \left[ \left( \frac{\gamma_1 \|z\|_2^{2p}}{2} + \frac{\|(\tilde{\Psi}_{T_{01}})^* \tilde{\Psi}_{T_{01}} z\|_2^p}{2\gamma_1} \right) + \rho^{2-p} \left( \frac{1}{c_1^{2p}} (1 + \gamma_2) \|z\|_2^{2p} + \left( 1 + \frac{1}{\gamma_2} \right) \eta^2 \right) \right].$$

Therefore

$$\left( 1 - \frac{c_2^p \gamma_1}{2} - \frac{\rho^{2-p} (1 + \gamma_2)}{c_1^{2p}} \right) \|z\|_2^{2p} \leq \frac{c_2^p}{2\gamma_1} \|(\tilde{\Psi}_{T_{01}})^* \tilde{\Psi}_{T_{01}} z\|_2^{2p} + \rho^{2-p} \left( 1 + \frac{1}{\gamma_2} \right) \eta^2$$

which in turn implies

$$\sqrt{\frac{2\gamma_1}{c_2^p}} \left[ \left( 1 - \frac{c_2^p \gamma_1}{2} - \frac{\rho^{2-p} (1 + \gamma_2)}{c_1^{2p}} \right) \|z\|_2^p - \rho^{1-p/2} \left( 1 + \frac{1}{\gamma_2} \right)^{1/2} \eta \right] \leq \|(\tilde{\Psi}_{T_{01}})^* \tilde{\Psi}_{T_{01}} z\|_2^p.$$

Using Proposition 3.6 *iii*) we conclude

$$\Gamma_1(p) \|z\|_2^p - \Gamma_2(p) \eta \leq (2\varepsilon)^p$$

where

$$\Gamma_1(p) = \sqrt{\frac{2\gamma_1}{c_2^p} \left( 1 - \frac{c_2^p \gamma_1}{2} - \frac{\rho^{2-p} (1 + \gamma_2)}{c_1^{2p}} \right) (1 - \delta_{s+M})^p} - \sqrt{(1 + \delta_M)^p \left( \frac{c_2}{c_1} \right)^p \rho^{2-p}}$$

and

$$\Gamma_2(p) = \sqrt{(1 - \delta_{s+M})^p \rho^{2-p} \left( 1 + \frac{1}{\gamma_2} \right) \frac{2\gamma_1}{c_1^p}} - \sqrt{\rho^{2-p} \left( \frac{c_2}{c_1} \right)^p (1 + \delta_M)^p}.$$

Thus choosing, for example,  $\gamma_1 \leq c_1^{-p}$  and  $M = s \left( \frac{2}{c_1^{2p}} \right)^{1/(2-p)}$  yields the result. Note that it is possible to choose the constants such that they obey

$$\Gamma'_1(p) < 0 \quad \text{and} \quad \Gamma'_2(p) > 0.$$

□

## 4 Numerics

The non-convex minimization problem  $(\ell^p\text{-P}_A^\varepsilon)$  is in general NP-hard although interior point methods exist to find local minimima, [23]. There are also other algorithms proposed to solve or rather approximate non-convex  $\ell^p$ -minimization problem and comparison of some methods can be found, for instance, in [40]. Many of these methods, are based on *iterative reweighting* which originally stems from [12] and has been used to strengthen the effect of sparsity. The ideas can be transferred to solve the analysis-based  $\ell^p$ -minimization problem. In the next section we present the algorithm that we use to solve  $(\ell^p\text{-P}_A^\varepsilon)$ . It can also be found in [21] for the synthesis formulation.

### 4.1 Approximating the $\ell^p$ -problem by reweighting

In order to solve the constrained  $\ell^p$ -analysis minimization problem  $(\ell^p\text{-P}_A^\varepsilon)$  under the presence of noise we use an algorithm that is based on the following intuitive discussion. Solving  $(\ell^p\text{-P}_A^\varepsilon)$ , that is solving

$$\min_x \|\Psi x\|_p^p \quad \text{s.t.} \quad \|y - Ax\|_2 \leq \varepsilon$$



is equivalent to solving

$$\min_x \sum_{\lambda} \frac{|\Psi x|_{\lambda}}{|\Psi x|_{\lambda}^{1-p}} \quad \text{s.t.} \quad \|y - Ax\|_2 \leq \varepsilon,$$

where we neglect for a moment the fact that dividing by  $|\Psi x|_{\lambda}$  is by no means always mathematically legit. Using the idea of reweighted  $\ell^1$ -minimization [12] we wish to strengthen the effect of sparsity by using the exact signal  $x_0$  to consider the weighted  $\ell^1$  problem

$$\min_x \sum_{\lambda} \frac{|\Psi x|_{\lambda}}{|\Psi x_0|_{\lambda}^{1-p}} \quad \text{s.t.} \quad \|y - Ax\|_2 \leq \varepsilon.$$

If  $(\Psi x_0)_{\lambda} = 0$ , the weight is set as  $\infty$ . Clearly, multiplying the denominator by some  $\mu > 0$  does not change the minimizer, hence, we consider

$$\min_x \sum_{\lambda} \frac{|\Psi x|_{\lambda}}{(\mu |\Psi x_0|_{\lambda})^{1-p}} \quad \text{s.t.} \quad \|y - Ax\|_2 \leq \varepsilon.$$

The purpose of  $\mu$  is to have better numerical control over the magnitude of the analysis coefficients and should be signal dependent. However, since  $x_0$  is not available in advance we might consider using an approximate vector  $x^k$  and consider

$$\min_x \sum_{\lambda} \frac{|\Psi x|_{\lambda}}{(\mu |\Psi x^k|_{\lambda})^{1-p}} \quad \text{s.t.} \quad \|y - Ax\|_2 \leq \varepsilon. \quad (4.1)$$

The approximate vector  $x^k$  can be obtained by solving (4.1) after an initialization of  $x_0$ . Finally, in order to prevent any instabilities one can introduce  $\nu > 0$  and solve

$$\min_x \sum_{\lambda} \frac{|\Psi x|_{\lambda}}{(\mu |\Psi x^k|_{\lambda} + \nu)^{1-p}} \quad \text{s.t.} \quad \|y - Ax\|_2 \leq \varepsilon. \quad (4.2)$$

The final program is then the following.

**Input** :  $y, \mu, \varepsilon, M$ .

**Output:**  $x$

Initialize  $k = 0, W = (w_{\lambda})_{\lambda} = 1$ .

**while**  $k \leq M$  **do**

Find solution of (4.2), i.e. find  $x^k$  such that

$$x^k = \operatorname{argmin}_x \|W \Psi x\|_1 \quad \text{s.t.} \quad \|y - Ax\|_2 \leq \varepsilon. \quad (4.3)$$

Update  $W$  by setting

$$W_{\lambda} = \frac{1}{(\mu |\Psi x^k|_{\lambda} + \nu)^{1-p}}.$$

Increase  $k \rightarrow k + 1$ .

**end**

**Algorithm 1:** Algorithm used to solve the minimization problem ( $\ell^p$ -P<sub>A</sub><sup>ε</sup>)

Note that (4.2) and (4.3), respectively, can be solved using many different common  $\ell^1$ -solvers such as NESTA ([3]) which is available at

<http://statweb.stanford.edu/~candes/nesta/>

## 4.2 Reconstruction from Fourier measurements using shearlets

Shearlets were first introduced in [26, 35] as a frame for the Hilbert space  $L^2(\mathbb{R}^2)$  that are based on *anisotropic scaling*, *shearing*, and *translations*. Its construction is not limited to the two dimensional case and the higher dimensional case has been considered in [32]. We shall not go into detail of the theoretical concepts of shearlets but refer the reader to [31, 29, 37, 33] for theoretical background and implementations of such systems.

The shearlet implementation used in this paper are downloaded from

<http://www.shearlab.org>

The object of interest is the GLPU phantom introduced in [25] which can be downloaded at

<http://bigwww.epfl.ch/algorithms/mriphantom/#soft>

As already mentioned in the introduction, one of the applications of the analysis-based  $\ell^p$ -minimization problem is magnetic resonance imaging where the sampling process is modeled by taking samples of the Fourier transform of the signal. The minimization problem for this application is

$$\min_x \|\Psi x\|_p^p \text{ s.t. } \|y - \mathcal{F}x\|_2 \leq \varepsilon \quad (4.4)$$

where  $\Psi$  is the shearlet transform and  $\mathcal{F}$  is the undersampled Fourier transform restricted onto a certain subset of frequencies. Typical choices for the set of measured frequencies are samples that lie on continuous trajectories such as horizontal lines, spirals, radial lines etc. For our numerical experiments we have chosen a radial sampling mask consisting of 30 radial lines, see Figure 3 below.

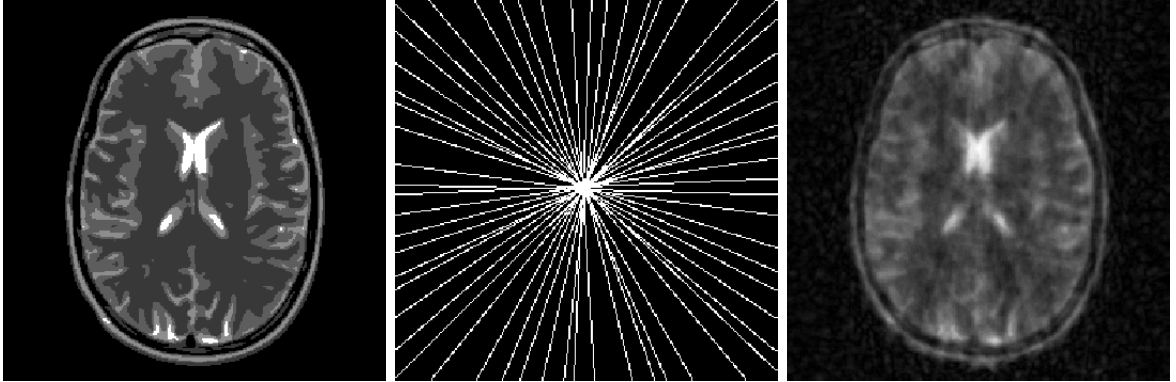


Figure 3: **Left:** GLPU Phantom from [25]. **Middle:** Radial sampling pattern used in Fourier domain. **Right:** Fourier inversion of data.

We now let Algorithm 1 run for a maximum of 10 iterations, i.e.  $M = 10$  in Algorithm 1. The outcome of the relative error is depicted in Figure 4. The curve for  $p = 1$  is not shown as it would be a constant line at the first error ( $\approx 17\%$ ). Otherwise, for decreasing  $p$  we see a decrease with every increasing iteration  $k$  of the error curves as  $p$  get smaller. It is important to mention, that for  $p = 1$  the result corresponds to conventional compressed sensing, i.e. standard  $\ell^1$ -minimization.

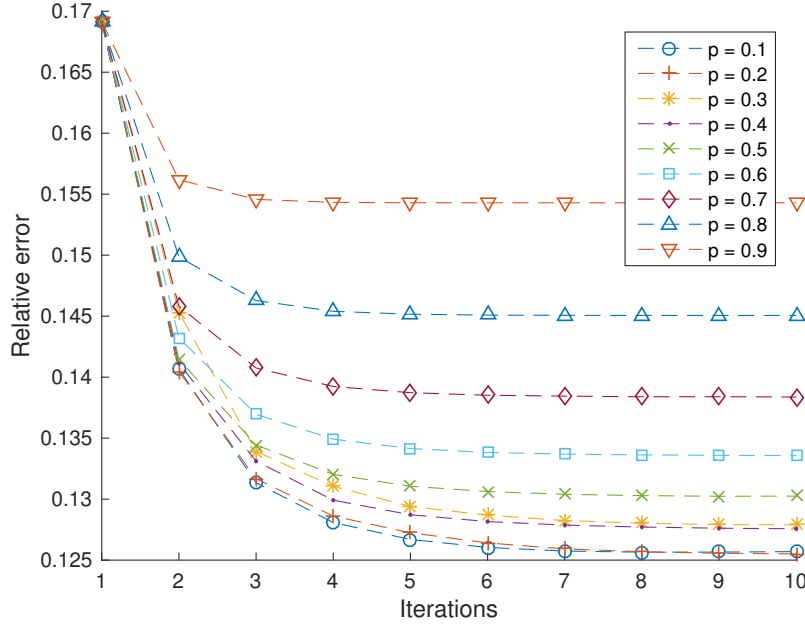


Figure 4: Relative error.

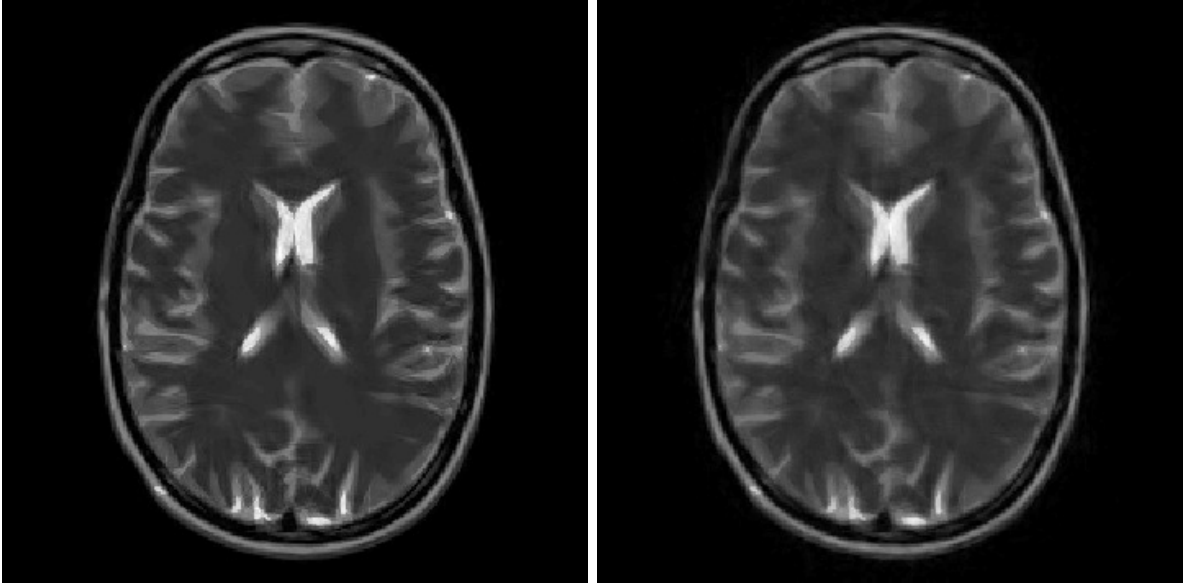


Figure 5: **Left:** Reconstruction for  $p = 0.1$ . **Right:** Reconstruction for  $p = 1$ .

Moreover, we like to mention that the other solutions are not obtained by using additional information except the computed sparsity structure that is known computed from the previous  $\ell^1$ -minimization. More precisely, the  $\ell^1$ -solution is used to obtain new weights which are now a good guess for the sparsity pattern. These weights are then returned to the algorithm so that it computes a new solution. In that sense the improvement is for free. However, It is computationally much more demanding as we have to solve the minimization problem  $M = 10$  times. Further, the algorithm takes longer as  $p$  gets smaller due to numerical instabilities, cf. Figure 6.

In Figure 6 we plot the relative error for a different number of iterations. Indeed, the first

$3 \times 3$  block plots show the nine sequences

$$\left( \frac{\|\Psi x_p^{k+1} - \Psi x_p^k\|}{\|\Psi x_p^{k+1}\|} \right)_{k=1, \dots, 10} \quad \text{for } p = 0.1, 0.2, \dots, 0.9,$$

where

$$x_p^{k+1} = \operatorname{argmin}_x \|W\Psi x\|_1 \quad \text{subject to} \quad \|y - Ax\|_2 \leq \varepsilon$$

with

$$W_\lambda = \frac{1}{(\mu|\Psi x^k|_\lambda + \nu)^{1-p}}.$$

That is, we show the outcome of Algorithm 1 for all  $p = 0.1, 0.2, \dots, 0.9$  and increasing  $k$  from 1 to 10. The second  $3 \times 3$  block then shows the same sequence but starting from  $k = 2$  for better visibility of the error curves

$$\left( \frac{\|\Psi x_p^{k+1} - \Psi x_p^k\|}{\|\Psi x_p^{k+1}\|} \right)_{k=2, \dots, 10} \quad \text{for } p = 0.1, 0.2, \dots, 0.9,$$

and the last one shows

$$\left( \frac{\|\Psi x_p^{k+1} - \Psi x_p^k\|}{\|\Psi x_p^{k+1}\|} \right)_{k=3, \dots, 10} \quad \text{for } p = 0.1, 0.2, \dots, 0.9.$$

By plotting the same sequence with  $k \geq 2$  and  $k \geq 3$  we can observe that the relative error decreases very quickly to zero after very few iterations which suggest fast numerical convergence of the algorithm. However, it also shows that such an approach via reweighting yields to less stable outcomes for very small  $p$  which is strongly visible in the third  $3 \times 3$  block plot for  $p = 0.1$  and larger  $k$ . The oscillations also show that the choice of a stopping criterion has to be done very carefully. We haven't incorporated any but the maximum number of iterations, which yield a termination if this number is reached.

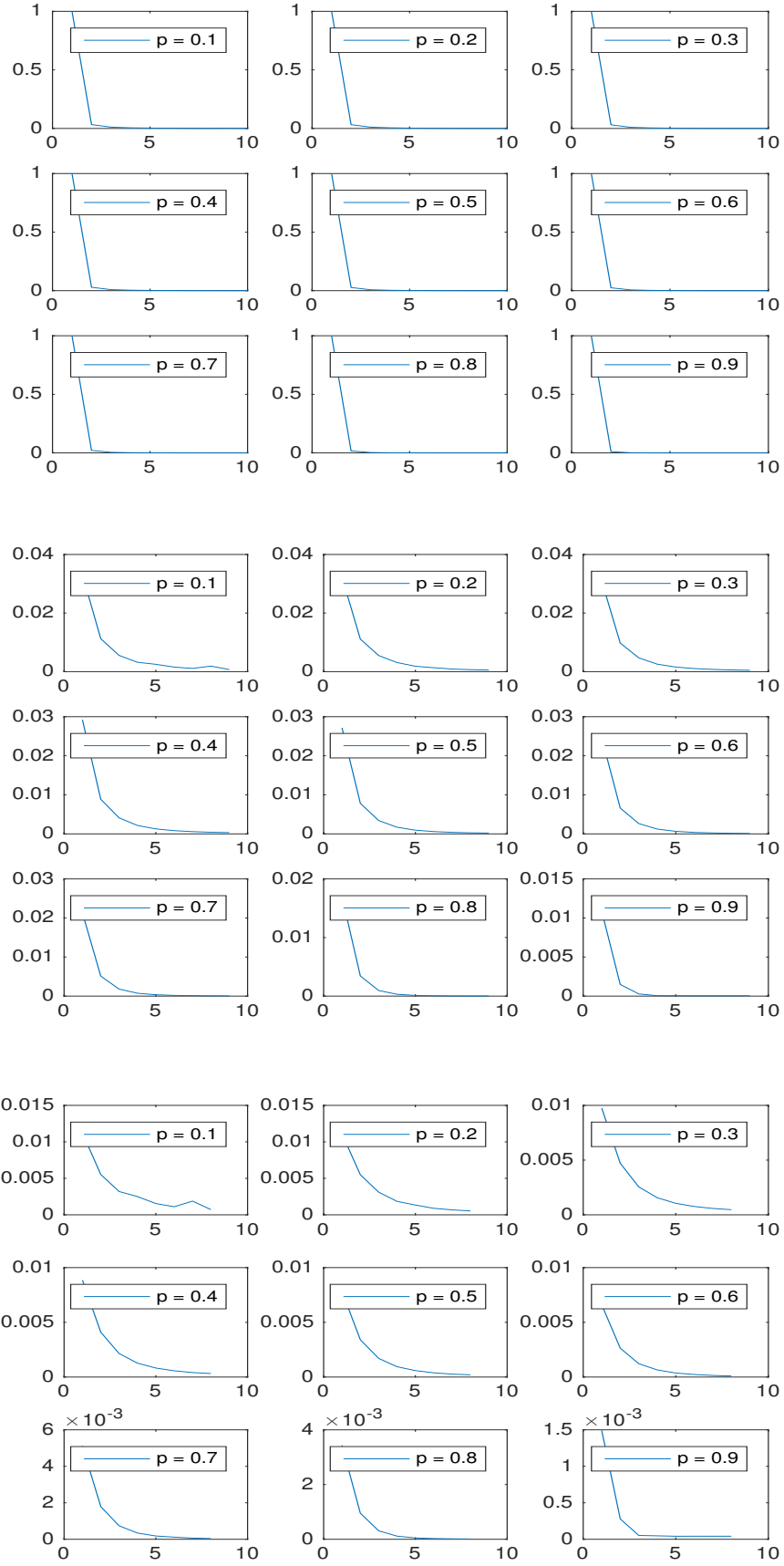


Figure 6: Relative error of the sequence of iterations. Starting from  $k = 1, 2$ , and  $3$ , respectively.

### 4.3 $q$ -controllability of shearlets

Finally we give a numerical outcome of the  $q$ -controllability of the shearlet frame bounds (2.6) in Table 1 as this is important in practice. In fact, the numbers show that one can choose  $q \approx 20000$  in this particular case. It is important to have this number  $q$  as large as possible since this in turn allows  $s$  to be large in Theorem 2.8.

| Scale $J$ | $c_1/c_2$ | $N^{-1}$ | $n \times n$       |
|-----------|-----------|----------|--------------------|
| 2         | 0.1152    | 7.266e-7 | $256 \times 256$   |
| 3         | 0.0889    | 3.913e-7 | $256 \times 256$   |
| 4         | 0.883     | 6.69e-8  | $512 \times 512$   |
| 5         | 0.0628    | 4.19e-8  | $512 \times 512$   |
| 6         | 0.0527    | 7.6e-9   | $1024 \times 1024$ |

Table 1: Verification of (2.6). In these cases  $q$  can be chosen to be  $\approx 1e6$ .

However, the numbers presented in Table 1 are very pessimistic and have to be interpreted in the right context as the discussion of the next section will show.

## 5 Redundancy versus sparsity in compressed sensing

In the previous content of this article we have only considered the analysis formulation for redundant transforms. This is theoretically necessary, otherwise if the transform corresponds to a basis there is no point in distinguishing between the analysis and synthesis formulation as these two problem would be equivalent. It has also been observed in applications that redundant transforms can yield better results. This is for example the case for the redundant wavelet transform in image restoration [46]. From that point of view redundancy greatly helps and one might argue that it is also needed or at least desired in certain applications. The purpose of this section is to argue that although redundancy seems to yield a great benefit, one has to be careful and discuss: How much redundancy is good in practice? Moreover, the redundancy factor should be a discussion on its own and should not be confused with the results of this paper.

We now consider two natural images of pixel size  $2848 \times 2848$ , shown in Figure 7 and demonstrate that typical (redundant) sparsifying transforms such as the wavelet and shearlet transform are from a compressed sensing point of view too redundant.

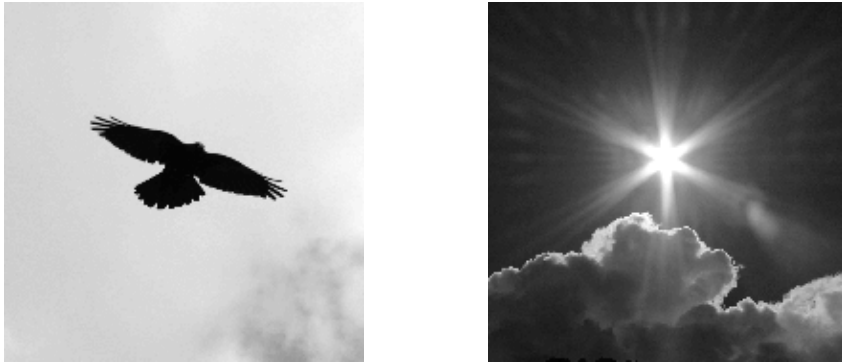


Figure 7: Reference images

### 5.1 Redundant wavelet transform

We first test the redundant wavelet transform that is made available by van den Berg and Friedlander in the `spot` toolbox which can be downloaded at

We first compute the wavelet decomposition of both reference images shown in Figure 7 for different total number of scales  $J = 2, 4, 6$ . Out of these wavelet coefficients we have computed the best  $s$ -term approximation using  $s = 90\%$  of the total number of pixels. This should not be confused with the number of total coefficients which is much larger. In the notation of our previous results we let  $s = 9/10 \cdot n$ , where  $n = 2848^2$  the dimension of the ambient space. The reconstructions are shown in Figure 8.



Figure 8: 90% of ambient dimension without subsampling (wavelets). Each column represents the reconstruction using 2, 4, and 6 scales respectively.

Obviously, the reconstructions get worse if more scales are available. More interestingly, the more coefficients there are available – obtained by increasing the number of scales – the more weight does the low frequency part gets in terms of the magnitudes of the coefficients resulting in an image that is very blurry. That shows that many more coefficients than the ambient dimension are needed in this case.

Now we conduct the same experiment again except that we do some subsampling before we take the best  $s$ -term approximation. More precisely, we use the same coefficients but only consider every forth wavelet coefficient and set all other coefficients to zero. Thus we divide the initial redundancy by a factor of 4. We then take the best  $s$ -term approximation using  $s = 30\%$  of the total number of pixels. The outcome can be seen in Figure 9. The reconstructed images in Figure 9 are significantly better than those obtained in Figure 8. From a theoretical point of view this behaviour is expected as the redundancy does not improve the approximation rate. However, it also hints that the intrinsic sparsity of the image in the analysis coefficients  $(\langle x, \psi_\lambda \rangle)_\lambda$  is much much smaller than what we can observe in the  $s$ -term approximation.

As we have already considered the shearlet transform in Section 4 of this paper we next want to show that this *curse of redundancy* can also be observed in that particular case.



Figure 9: 30% of ambient dimension with subsampling (wavelets). Each column represents the reconstruction using 2, 4, and 6 scales respectively.

## 5.2 Redundant shearlet transform

In this section we show the same experiment but with shearlets instead of wavelets. It is important to mention that the shearlet transform is truly redundant in the sense that there exists no non-redundant shearlet transform in the literature so far.



Figure 10: 90% of ambient dimension without subsampling (shearlets) using 2, 4, and 6 scales.



In Figure 10 we show the best  $s$ -term approximation using again 90% of the ambient dimension for a 2, 4, and 6, level shearlet decomposition, respectively. We again observe that the images get more blurry in Figure 8. Similar as for wavelets we can observe that the low frequency part gets more and more important if more scales are activated.

Now we again subsample the already computed shearlet coefficients by only considering every forth coefficient and delete all others by setting them to zero. The reconstructions are shown in Figure 11. Again, one can observe a significant improvement using only 10% of the ambient dimension. Note that by increasing the scales and in that way adding more elements to the dictionary the system gets more and more redundant and the reconstructions get more and more accurate. This may sound confusing as we observed the opposite behaviour in Figure 10. There we saw that the redundancy made the  $s$ -term approximation worse.



Figure 11: 10% of ambient dimension with subsampling (shearlets) using 2, 4, and 6 scales.

It is very evident from both numerical experiments that there is an intrinsic sparsity contained in the analysis coefficients of a redundant transform that is not fully characterized by the sparsity assumption alone. One possible approach to tackle this problem from a different perspective could be to rely on the *statistical dimension*, [2]. This is, however, not part of this work. Further, note that it is not correct to say, that the compression rate is not strong enough to make the results of this paper to work as Figure 11 shows that clearly if one were to use a subsampled shearlet transform then the compression rate is good enough for our results to apply.

## 6 Discussion and future work

As we discussed in Section 2, it is the authors believe that the minimization problem should be performed over dual coefficients instead of transform coefficients if the  $\Psi$ -RIP is to be assumed. Surely, if the dual system and the primal system give rise to the same sparsity pattern than this argument should no longer be valid which was in our analysis demonstrated by the concept of identifiable duals. However, the design of sparse representation systems that have good duals

is a challenging task in applied harmonic analysis. Nevertheless, it is necessary if one wants to combine compressed sensing with redundant dictionaries.

Furthermore, we want to comment on some problems that are left for future work:

- In order to find the optimal  $p$  the constants  $C_1(p), C_2(p)$  should be optimized. For this one has to optimize over  $p, \gamma_1, \gamma_2, \delta, M, s$ .
- The class of frames that have an identifiable dual can be seen as a generalization of scalable frames. Since the latter have nice geometric characterizations [34] via open quadratic cones, it is interesting to see whether similar characterizations can be computed for frames that have an identifiable dual.
- Another question left for future work is the replacement of the  $\Psi$ -RIP by an isometry condition as it has been done for the synthesis model in [7] in order to have a RIPless theory for the analysis formulation of the  $\ell^p$ -minimization problem.
- Also an infinite dimensional scenario could be investigated for the  $\ell^p$  scenario, in particular, in this case the  $q$ -controllability(2.6) should be replaced by another condition that may shed more light on the redundancy problem.
- The redundancy is a very subtle issue. In Section 5 we have seen in Figure 9 and 2, respectively, that there is an intrinsic sparsity structure that has not been captured by the theory yet. It also shows that sparsity alone as in the classical sense, is not sophisticated enough to explain why the analysis formulation works. In particular, the subsampling issue is coming from the implementations of such transforms, in our cases wavelets and shearlets, not from the actual theory. However, a possible point of future work is to mathematically quantify this redundancy that arises in the discrete setting as presented in Section 5. A possible approach is to involve the statistical dimension or other geometric properties of the problem, [2, 48].

## Acknowledgements

The author would like to thank Ben Adcock for inspiring discussions and acknowledges support from the Berlin Mathematical School as well as the DFG Collaborative Research Center TRR 109 "Discretization in Geometry and Dynamics".

## A Proof of Theorem 2.6

The optimal  $\delta_s$  for which  $\tilde{A} = \begin{pmatrix} r_1 \\ \vdots \\ r_m \end{pmatrix} \in \mathbb{R}^{m \times n}$  satisfies the  $\Psi$ -RIP is given by

$$\delta_s = \|\tilde{A}^* \tilde{A} - \text{Id}_n\|_{\Delta} := \sup_{f \in \Delta} \langle (\tilde{A}^* \tilde{A} - \text{Id}_n)x, x \rangle$$

where

$$\Delta = \{x \in \text{ran } \Psi^* : x = \Psi^* c, \|c\|_0 \leq s, \|x\| \leq 1\} \subseteq \mathbb{R}^n.$$

Since  $\mathbb{E} r_i^* r_i = \text{Id}_n$  we have

$$\delta_s = \|\tilde{A}^* \tilde{A} - \text{Id}_n\|_{\Delta} = \left\| \frac{1}{m} \sum_{i=1}^m \frac{1}{m_i} r_i^* r_i - \text{Id}_n \right\|_{\Delta} = \frac{1}{m} \left\| \sum_{i=1}^m (r_i^* r_i - \mathbb{E} r_i^* r_i) \right\|_{\Delta}.$$

For a Rademacher sequence  $\varepsilon = (\varepsilon_i)_i$  independent of  $(r_i)_i$  we have by Lemma 6.7 in [43]

$$\mathbb{E}\delta_s = \frac{1}{m}\mathbb{E}\left\|\sum_{i=1}^m(r_i^*r_i - \mathbb{E}r_i^*r_i)\right\|_{\Delta} \leq \frac{2}{m}\mathbb{E}\left\|\sum_{i=1}^m\varepsilon_ir_i^*r_i\right\|_{\Delta},$$

hence,

$$\mathbb{E}\delta_s \leq \frac{2}{m}\mathbb{E}_r\mathbb{E}_{\varepsilon}\sup_{x\in\Delta}\left|\left\langle\sum_{i=1}^m\varepsilon_ir_i^*r_ix, x\right\rangle\right| = \frac{2}{m}\mathbb{E}_r\mathbb{E}_{\varepsilon}\sup_{x\in\Delta}\left|\sum_{i=1}^m\varepsilon_i|\langle r_i, x\rangle|^2\right|.$$

Define the pseudo-metric

$$d(x, y) = \left(\sum_{i=1}^m(|\langle r_i, x\rangle|^2 - |\langle r_i, y\rangle|^2)^2\right)^{1/2}.$$

Then, as shown in [30] we have for  $x, y \in \Delta$

$$d(x, y) \leq 2\sup_{z\in\Delta}\left(\sum_{i=1}^m|\langle r_i, z\rangle|^{2p}\right)^{1/(2p)}\left(\sum_{i=1}^m|\langle r_i, x-y\rangle|^{2p}\right)^{1/(2p)},$$

where  $p, q \geq 1$  such that  $p^{-1} + q^{-1} = 1$ .

Now, for any  $h \in \Delta$  of the form  $z = \Psi^*c$  with  $\|c\|_0 \leq s$  and any realization of  $(r_i)_i$  we have

$$\begin{aligned} |\langle r_i, z\rangle| &= |\langle \Psi(\Psi^*\Psi)^{-1}r_i, \Psi\Psi^*c\rangle| \\ &\leq \sum_{\lambda \leq N} |\langle r_i, \tilde{\psi}_{\lambda}\rangle| |(\Psi\Psi^*c)_{\lambda}| \\ &\leq \sum_{\lambda \leq N} \frac{K}{c_1} |(\Psi\Psi^*c)_{\lambda}|, \end{aligned}$$

where  $K > 0$  is so that  $\|\psi_{\lambda}\| \leq K$  for all  $\lambda \leq N$  and  $c_1$  denotes the lower frame bound. Therefore

$$|\langle r_i, h\rangle| \leq \frac{K}{c_1}L\sqrt{s}$$

with

$$L = \sup_{\substack{\|\Psi^*c\|=1 \\ \|c\|_0 \leq s}} \frac{\|(\Psi\Psi^*c)_{\lambda}\|_1}{\sqrt{s}}.$$

Therefore we obtain

$$\begin{aligned} \sup_{z\in\Delta}\left(\sum_{i=1}^m|\langle r_i, z\rangle|^{2p}\right)^{1/(2p)} &= \sup_{z\in\Delta}\left(\sum_{i=1}^m|\langle r_i, z\rangle|^2|\langle r_i, z\rangle|^{2p-2}\right)^{1/(2p)} \\ &\leq \left(s\left(\frac{KL}{c_1}\right)^2\right)^{(p-1)/(2p)}\left(\sup_{z\in\Delta}\sum_{i=1}^m|\langle r_i, z\rangle|^{2p}\right)^{1/(2p)}. \end{aligned}$$

The rest of the proof follows the argumentation given in [30].

For a set  $\Sigma$ , a metric  $d$  and a given  $t > 0$  the *covering number*  $\mathcal{N}(\Sigma, d, t)$  is defined as the smallest number of balls of radius  $t$  centered at points of  $\Sigma$  necessary to cover  $\Sigma$  with respect to  $d$ . By Dudley's inequality we have

$$\mathbb{E}_{\varepsilon}\sup_{x\in\Delta}\left|\left\langle\sum_{i=1}^m\varepsilon_ir_i^*r_ix, x\right\rangle\right| \leq 4\sqrt{2}\int_0^{\infty}\sqrt{\log(\mathcal{N}(\Delta, d, t))}dt. \quad (\text{A.1})$$

Using the semi-norm

$$\|x\|_{X,q} := \left( \sum_{i=1}^m |\langle r_i, x \rangle|^{2q} \right)^{1/(2q)}$$

we obtain using covering arguments and (A.1)

$$\begin{aligned} \mathbb{E}_\epsilon \sup_{x \in \Delta} \left| \left\langle \sum_{i=1}^m \varepsilon_i r_i^* r_i x, x \right\rangle \right| \\ \leq C \left( s \left( \frac{KL}{c_1} \right)^2 \right)^{(p-1)/(2p)} \left( \sum_{i=1}^m |\langle r_i, x \rangle|^{2q} \right)^{1/(2q)} \int_0^\infty \sqrt{\log(\mathcal{N}(\Delta, \|\cdot\|_{X,q}, t))} dt. \end{aligned}$$

Now, following the arguments in [30] we have

$$\int_0^\infty \sqrt{\log(\mathcal{N}(\Delta, \|\cdot\|_{X,q}, t))} dt \leq C \sqrt{q(sL^2)m^{1/q} \log(n) \log^2(sL^2)}.$$

Thus in (A.1) we obtain

$$\begin{aligned} \mathbb{E}\delta_s &\leq \frac{C \left( s \left( \frac{KL}{c_1} \right)^2 \right)^{(p-1)/(2p)} \sqrt{qm^{1/q} sL^2 \log(n) \log^2(sL^2)}}{m} \mathbb{E} \sup_{x \in \Delta} \left( \sum_{i=1}^m |\langle r_i, x \rangle|^2 \right)^{1/(2p)} \\ &\leq \frac{C \left( s \left( \frac{KL}{c_1} \right)^2 \right)^{(p-1)/(2p)} \sqrt{q \log(n) \log^2(sL^2)}}{m^{1-1/(2q)-1/(2p)}} \mathbb{E} \left( \frac{1}{m} \left\| \sum_{i=1}^m r_i^* r_i - \text{Id}_n \right\|_\Delta + \left\| \text{Id}_n \right\|_\Delta \right)^{1/(2p)} \\ &\leq \frac{C \left( s \left( \frac{KL}{c_1} \right)^2 \right)^{(p-1)/(2p)} \sqrt{q \log(n) \log^2(sL^2)}}{m^{1/2}} \sqrt{\mathbb{E}\delta_s + 1}. \end{aligned}$$

We can assume  $K/c_1$  to be greater than one, hence

$$\mathbb{E}\delta_s \leq \frac{C \left( s \left( \frac{KL}{c_1} \right)^2 \right)^{(p-1)/(2p)} \sqrt{q \log(n) \log^2(s \left( \frac{KL}{c_1} \right)^2)}}{m^{1/2}} \sqrt{\mathbb{E}\delta_s + 1}.$$

Choosing  $p = 1 + (\log(s(KL)^2 c_1^{-2}))^{-1}$  and  $q = 1 + \log(s(KL)^2 c_1^{-2})$  yields

$$(s(KL)^2 c_1^{-2})^{1/2+(p-1)/(2p)} \leq \sqrt{e},$$

hence,

$$\mathbb{E}\delta_s \leq C \sqrt{2 \log(n) \log^2(s(KL)^2 c_1^{-2})/m} \sqrt{\mathbb{E}\delta_s + 1}.$$

Finally,

$$\mathbb{E}\delta_s \leq C \sqrt{\frac{s(KL)^2 c_1^{-2} \log(n) \log^3(sL^2)}{m}}$$

provided  $\frac{s(KL)^2 c_1^{-2} \log(n) \log^3(s(KL)^2 c_1^{-2})}{m} \leq 1$ . Therefore,  $\mathbb{E}\delta_s \leq \delta/2$  for some  $\delta \in (0, 1)$  if

$$m \geq C \delta^{-2} s(KL)^2 c_1^{-2} \log^3(s(KL)^2 c_1^{-2}) \log N. \quad (\text{A.2})$$

Let  $f_{x,y}(r) = \text{Re}(\langle (r_i^* r_i - \text{Id}_n)z, w \rangle)$  so that

$$m\delta_s = \left\| \sum_{i=1}^m (r_i^* r_i - \mathbb{E} r_i^* r_i) \right\|_{\Delta} = \sup_{x,y \in \Delta} \sum_{i=1}^m f_{x,y}(r_i).$$

Note that we have

- ▷  $\mathbb{E} f_{x,y}(r_i) = 0$ ,
- ▷  $|f_{x,y}(r)| \leq s(KL)^2 c_1^{-2} + 1$ ,
- ▷  $\mathbb{E} |f_{x,y}(r)|^2 = \mathbb{E} \|(r_i^* r_i - \text{Id})x\|_2^2 \leq (s(KL)^2 c_1^{-2} + 1)^2$ .

Now, fix some  $\delta \in (0, 1)$  and choose  $m$  in accordance with (A.2). Then by Theorem 6.25 of [43] we have

$$\begin{aligned} \mathbb{P}(\delta_s \geq \delta) &\leq \mathbb{P}(\delta_s \geq \mathbb{E}\delta_s + \delta/9) \\ &= \mathbb{P}\left(\left\| \sum_{i=1}^m (r_i^* r_i - \mathbb{E} r_i^* r_i) \right\|_{\Delta} \geq \mathbb{E} \left\| \sum_{i=1}^m (r_i^* r_i - \mathbb{E} r_i^* r_i) \right\|_{\Delta} + \delta m/9\right) \\ &\leq \exp\left(-\frac{\left(\frac{\delta m}{9(s(KL)^2 c_1^{-2} + 1)}\right)^2}{2m \left(1 + \frac{\delta}{s(KL)^2 c_1^{-2} + 1}\right) + \frac{2}{3} \left(\frac{\delta m}{9(s(KL)^2 c_1^{-2} + 1)}\right)}\right) \\ &\leq \exp\left(-\frac{\delta^2 m}{Cs(KL)^2 c_1^{-2}}\right), \end{aligned} \tag{A.3}$$

where the constant  $C$  might have changed in the last estimate. Further, if

$$m \geq C\delta^{-2} s(KL)^2 c_1^{-2} \log(1/\gamma),$$

then (A.3) is bounded by  $\gamma$ . Thus,  $\delta_s \leq \delta$  with probability  $1 - \gamma$  if

$$m \geq C\delta^{-2} s(KL)^2 c_1^{-2} \max\{\log^3(s(KL)^2 c_1^{-2}) \log(N), \log(1/\gamma)\}.$$

The proof is complete.

## References

- [1] A. Aldroubi, X. Chen, and A. M. Powell. Perturbations of measurement matrices and dictionaries in compressed sensing. *Appl. Comput. Harmon. Anal.*, 33(2):282–291, 2012.
- [2] D. Amelunxen, M. Lotz, M. B. McCoy, and J. A. Tropp. *Inform. Inference*, 15(3):224–294, 2014.
- [3] S. Becker, J. Bobin, and E. J. Candès. NESTA: A fast and accurate first-order method for sparse recovery. *SIAM J. Imaging Sci.*, 4(1):1–39, 2011.
- [4] E. J. Candès. The restricted isometry property and its implications for compressed sensing. *C. R. Math. Acad. Sci. Paris*, 346(9-10):589–592, 2008.
- [5] E. J. Candès and D. L. Donoho. New Tight Frames of Curvelets and Optimal Representations of Objects with piecewise  $C^2$ -Singularities. *Comm. Pure Appl. Math.*, pages 219–266, 2002.

- [6] E. J. Candès, Y. C. Eldar, D. Needell, and P. Randall. Compressed sensing with coherent and redundant dictionaries. *Appl. Comput. Harmon. Anal.*, 31(1):59–73, 2011.
- [7] E. J. Candès and Y. Plan. A Probabilistic and RIPless Theory of Compressed Sensing. *IEEE Trans. Inform. Theory*, 57(11):7235–7254, 2011.
- [8] E. J. Candès, J. K. Romberg, and T. Tao. Stable signal recovery from incomplete and inaccurate measurements. *Comm. Pure Appl. Math.*, 59:1207–1223, 2005.
- [9] E. J. Candès, J. K. Romberg, and T. Tao. Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information. *IEEE Trans. Inform. Theory*, 52(2):489–509, 2006.
- [10] E. J. Candès and T. Tao. Decoding by linear programming. *IEEE Trans. Inform. Theory*, 51(12):4203–4215, 2005.
- [11] E. J. Candès and T. Tao. Near-optimal signal recovery from random projections: universal encoding strategies? *IEEE Trans. Inform. Theory*, 52(12):5406–5425, 2006.
- [12] E. J. Candès, M. B. Wakin, and S. P. Boyd. Enhancing sparsity by reweighted  $\ell_1$  minimization. *J. Fourier Anal. Appl.*, 14(5-6):877–905, 2008.
- [13] R. Chartrand. Exact reconstruction of sparse signals via nonconvex minimization. *IEEE Signal Process. Lett.*, 14:707–710, 2007.
- [14] O. Christensen. *An Introduction to Frames and Riesz Bases*. Applied and Numerical Harmonic Analysis. Birkhäuser Boston, Inc., Boston, MA, 2003.
- [15] A. Cohen, W. Dahmen, and R. DeVore. Compressed sensing and best  $k$ -term approximation. *J. Amer. Math. Soc.*, pages 211–231, 2009.
- [16] I. Daubechies. *Ten Lectures on Wavelets*, volume 61 of *CBMS-NSF Regional Conference Series in Applied Mathematics*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 1992.
- [17] D. L. Donoho. Compressed sensing. *IEEE Trans. Inform. Theory*, 52:1289–1306, 2006.
- [18] M. Elad, P. Milanfar, and R. Rubinstein. In *Signal Processing Conference, 2006 14th European*, pages 1–5, 2006.
- [19] M. Fornasier and K. Gröchenig. Intrinsic localization of frames. *Constr. Approx.*, 22(3):395–415, 2005.
- [20] S. Foucart. A note on guaranteed sparse recovery via  $\ell_1$ -minimization. *Appl. Comput. Harmon. Anal.*, 29(1):97–103, 2010.
- [21] S. Foucart and M.-J. Lai. Sparsest solutions of underdetermined linear systems via  $l_q$ -minimization for  $0 < q \leq 1$ . *Appl. Comput. Harmon. Anal.*, 26(3):395–407, 2009.
- [22] S. Foucart and H. Rauhut. *A Mathematical Introduction to Compressive Sensing*. Birkhäuser Basel, 2013.
- [23] D. Ge, X. Jiang, and Y. Ye. A note on the complexity of  $L_p$  minimization. *Math. Program.*, 129(2):285–299, 2011.
- [24] P. Grohs. Intrinsic Localization of Anisotropic Frames. *Appl. Comput. Harmon. Anal.*, 35(2):264 – 283, 2013.

- [25] M. Guerquin-Kern, L. Lejeune, K. P. Pruessmann, and M. Unser. Realistic Analytical Phantoms for Parallel Magnetic Resonance Imaging. *IEEE Trans. Med. Imaging*, 31(3):626–636, 2012.
- [26] K. Guo, G. Kutyniok, and D. Labate. Sparse Multidimensional Representations using Anisotropic Dilation and Shear Operators. *Wavelets and splines: Athens 2005*, 1:189–201, 2006.
- [27] M. Kabanava and H. Rauhut. Analysis  $\ell_1$ -recovery with frames and gaussian measurements. *Acta Appl. Math.*, 140(1):173–195, 2015.
- [28] M. Kabanava, H. Rauhut, and H. Zhang. Robust analysis  $\ell_1$ -recovery from gaussian measurements and total variation minimization. *European J. Appl. Math.*, 26(6):917–929, 2015.
- [29] P. Kittipoom, G. Kutyniok, and W.-Q Lim. Construction of Compactly Supported Shearlet Frames. *Constr. Approx.*, 35(1):21–72, 2012.
- [30] F. Krahmer, D. Needell, and R. Ward. Compressive sensing with redundant dictionaries and structured measurements. *SIAM J. Math. Anal.*, to appear., 2015.
- [31] G. Kutyniok and D. Labate. Introduction to Shearlets. In *Shearlets*, Appl. Numer. Harmon. Anal., pages 1–38. Birkhäuser/Springer, New York, 2012.
- [32] G. Kutyniok, J. Lemvig, and W.-Q. Lim. Optimally sparse approximations of 3D functions by compactly supported shearlet frames. *SIAM J. Math. Anal.*, 44(4):2962–3017, 2012.
- [33] G. Kutyniok, W.-Q Lim, and R. Reisenhofer. ShearLab 3D: Faithful digital shearlet transforms based on compactly supported shearlets. *ACM Trans. Math. Software*, 42(1), 2016.
- [34] G. Kutyniok, K. A. Okoudjou, F. Philipp, and E. K. Tuley. Scalable frames. *Linear Algebra Appl.*, 438(5):2225–2238, 2013.
- [35] D. Labate, W.-Q Lim, G. Kutyniok, and G. Weiss. Sparse multidimensional representation using shearlets. *Wavelets XI., Proceedings of the SPIE*, 5914:254–262, 2005.
- [36] W.-Q Lim. The discrete shearlet transform: a new directional transform and compactly supported shearlet frames. *IEEE Trans. Image Process.*, 19(5):1166–1180, 2010.
- [37] W.-Q Lim. Nonseparable shearlet transform. *IEEE Trans. Image Process.*, 22(5):2056–2065, 2013.
- [38] Y. Liu, T. Mi, and S. Li. Compressed sensing with general frames via optimal-dual-based  $\ell^1$ -analysis. *IEEE Trans. Inform. Theory*, 58(7):4201–4214, 2012.
- [39] M. Lustig, D. Donoho, and J. Pauly. Sparse MRI: The application of compressed sensing for rapid MR imaging. *Magn. Reson. Med.*, 58(6):1182–1195, 2007.
- [40] Q. Lyu, Z. Lin, Y. She, and C. Zhang. A comparison of typical  $\ell_p$  minimization algorithms. *Neurocomputing*, 119(0):413 – 424, 2013.
- [41] D. Mücke-Herzberg, P. Abellan, M. C. Sarahan, I. S. Godfrey, Z. Saghi, R. K. Leary, A. Stevens, J. Ma, G. Kutyniok, F. Azough, R. Freer, P. A. Midgley, N. D. Browning, and Q. M. Ramasse. Practical Implementation of Compressive Sensing for High Resolution STEM. *Microscopy and Microanalysis*, 22:558–559, 2016.
- [42] S. Nam, M. Davies, M. Elad, and R. Gribonval. The cosparsity analysis model and algorithms. *Appl. Comput. Harmon. Anal.*, 34(1):30–56, 2013.

- [43] H. Rauhut. Compressive sensing and structured random matrices. In M. Fornasier, editor, *Theoretical Foundations and Numerical Methods for Sparse Recovery*, volume 9 of *Radon Series Comp. Appl. Math.*, pages 1–92. deGruyter, 2010.
- [44] R. Reisenhofer, J. Kiefer, and E. J. King. Shearlet-based detection of flame fronts. *Exp. Fluids.*, 57(3):1–14, 2016.
- [45] R. Saab, R. Chartrand, and Ö. Yilmaz. Stable sparse approximations via nonconvex optimization. In *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing, ICASSP 2008*, pages 3885–3888, 2008.
- [46] J.-L. Starck, F. Murtagh, and J. Fadili. *Sparse Image and Signal Processing: Wavelets and Related Geometric Multiscale Analysis*. Cambridge University Press, 2015.
- [47] S. Vaiter, G. Peyre, C. Dossal, and J. Fadili. Robust sparse analysis regularization. *IEEE Trans. Inform. Theory*, 59(4):2001–2016, 2013.
- [48] R. Vershynin. *Estimation in high dimensions: a geometric perspective*. Birkhauser Basel, 2015.